



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Social interaction in virtual environments:  
The relationship between mutual gaze, task  
performance and social presence.**

*Sara Dalzel-Job*



Doctor of Philosophy

Institute of Communicative and Collaborative Systems

School of Informatics

University of Edinburgh

2014

# Abstract

Everyday face-to-face social interaction is increasingly being supplemented by computer- and video-mediated communication. With mediation, however, comes the potential loss of important non-verbal cues. It is therefore important to attempt to maintain the quality of the mediated interaction, such that it retains as many of the aspects of a real-world interaction as possible. Social presence is a measure of how similar a mediated interaction is to face-to-face, the most socially present situation, in terms of perceptions of and behaviour towards an interlocutor. Social presence can be mediated by many factors, one of which is mutual gaze, and social perceptions of an interlocutor are also thought to be related to task performance. For a successful interaction, therefore, an optimum amount of mutual gaze for maximising social presence and task performance is desirable. This research aims to investigate the relationship between mutual gaze, task performance and social presence, in order to discover the ideal conditions under which a successful mediated interaction can occur.

Previous gaze research paradigms have involved one conversational partner staring continuously at the other, and the resulting mutual gaze being measured. It is hypothesised that this method may actually suppress mutual gaze, primarily due to social reasons. It is potentially, therefore, not the optimum experimental design for mutual gaze research. The first study in this thesis used eye-tracking to explore this hypothesis and investigate the relationship between mutual gaze and task performance. A suitable paradigm was developed, based on that used in previous research into eye movements and non-verbal communication. Two participants – Instruction Giver (IG) and Instruction Follower (IF) – communicated via avatars in Second Life to solve simple arithmetic tasks. There were two between-participant *looking conditions*: staring (the IG's avatar stared continuously at the IF); and not-staring, (IG's avatar looked at IF and task-relevant objects). Constant staring did, indeed, show evidence of decreasing mutual gaze within the dyad. Mutual gaze was positively correlated with task performance scores, but only in the not-staring condition. When not engaged in mutual gaze, the IF looked more at task-related objects in the not-staring condition than in the staring condition; this suggests that

social factors are likely to be driving the gaze aversion in the staring condition. Furthermore, there are no task-related benefits to staring.

The second study explored further how much looking by one person at another will maximise both mutual gaze and task performance between the dyad. It also investigated the relationship between mutual gaze, task performance and both manipulated and perceived social presence. Individual participants interacted with a virtual agent within the Second Life paradigm previously used in the human-human study. Participants were either told they were interacting with a computer (i.e. an agent) or another human (an avatar). This provided the between-participants manipulated social presence variable, or *agency*. The virtual agent was programmed to look at the participant during either 0%, 25%, 50% or 75% of the interaction, providing the within-participants variable *looking condition*. The majority of effects were found in the 75% looking condition, including the highest mutual gaze uptake and the highest social presence ratings (measured via a questionnaire). Although the questionnaire did not detect any differences in social presence between the agent and avatar condition, participants were significantly faster to complete the tasks in the avatar condition than in the agent condition. This suggests that behavioural measures may be more effective at detecting differences in social presence than questionnaires alone. The results are discussed in relation to different theories of social interaction. Implications and limitations of the findings are considered and suggestions for future work are made.

## **Lay summary**

Our society relies, ever-increasingly, on interacting with people using computers (computer-mediated communication, or CMC) Although meetings via Skype, online gaming and teaching or business meetings using virtual environments can be more practical than face-to-face interactions, important social cues, such as eye contact, can be lost. Social presence is a measure of how similar a mediated interaction is to a face-to-face one, in terms of perceptions of, and behaviour towards, a conversational partner. Social presence perceptions are thought to be related to access to eye contact (or mutual gaze) between a mediated pair of individuals, as well as impacting on how successful an interaction is, specifically in terms of any joint task that is being carried out.

Previous investigations into eye contact have involved one person staring continuously at another and the resulting mutual gaze being measured. This thesis suggests that this is not the most appropriate method for measuring mutual gaze, as the act of staring at someone may, itself, make them less likely to want to make eye contact. The first experiment in this thesis investigates this hypothesis. It looks at the relationship between eye contact and task performance during a two-person task within the virtual environment, Second Life. Two participants – Instruction Giver (IG) and Instruction Follower (IF) – communicated via avatars in Second Life to solve simple arithmetic tasks. One group of participants had the IG's avatar staring continuously at the IF, and in the other group, the IG's avatar looked at the IF intermittently. It was found that staring did indeed reduce the amount of mutual gaze between the pair, as well as reducing task performance. In the not-staring group, however, more mutual gaze was related to better task performance.

The second study looked at how much looking by one person at another will maximise both mutual gaze and task performance. It also investigated the relationship between mutual gaze, task performance and both manipulated and perceived social presence. Individuals worked with a virtual agent IG within the Second Life paradigm previously used in the human-human study. Participants were either told they were interacting with a computer (i.e. an agent) or another human (an avatar). The virtual agent was programmed to look at the participant for varying

amounts of time during the interaction. The highest amount of mutual gaze and the highest social presence ratings (measured by a questionnaire) were found when the IG looked at the IF during 75% of the interaction. Although the questionnaire did not detect any differences in social presence between the agent and avatar condition, tasks were completed faster in the avatar condition than in the agent condition. This suggests that behavioural measures such as task performance, may be more effective at detecting differences in social presence than questionnaires alone. The results are discussed in relation to different theories of social interaction.

## Acknowledgements

I would like to thank my supervisors, Jon Oberlander and Tim Smith for their unending encouragement, guidance and patience, and for maintaining a sense of humour throughout. Many thanks to Craig Nicol, whose help and persistence with the set-up of the original paradigm, along with a fair amount of trouble-shooting, was invaluable. I am also grateful to Jeff Dalton and Ed Baggs for all their help with programming Tracker. Thanks to Austin Tate for all his assistance with Second Life, and to Robin Hill, who was not only extremely helpful with all things eye tracking, but also a reassuring and cheerful face.

My family have all been incredibly supportive throughout this process and I would particularly like to thank my parents for their unwavering practical, emotional and financial assistance. I would also like to mention my son, Mylo, whose two-year-old chatter provided a marvellous counterpoint to statistics. Finally, I am incredibly grateful to my husband Eliot, who has been behind me all the way, and without whom I would not have been able to complete this PhD.

## Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Work from published papers under joint authorship appears as follows:

**Chapters 2 and 3:** Aspects of the human-human study and some of the supporting literature and discussion, as well as an outline for the human-agent study was published in ISPR (2011). All work is credited to the candidate; supervisory input was nominal:

Dalzel-Job, S., Oberlander, J., & Smith, T. J. (2011). Contested staring: issues and the use of mutual gaze as an on-line measure of social presence. Paper presented at the *2011 International Society for Presence Research Annual Conference*, Edinburgh, UK.

**Chapter 2:** The mutual gaze – task performance results from the human-human experiment were published in CogSci (2011). All work is credited to the candidate; supervisory input was nominal:



Dalzel-Job, S., Oberlander, J., & Smith, T. J. (2011). Don't Look Now: The relationship between mutual gaze, task performance and staring in Second Life. Paper presented at the *The 33<sup>rd</sup> Annual Conference of the Cognitive Science*, Boston, Massachusetts, USA, 20-23 July 2011.



# Table of Contents

Abstract .....	ii
Acknowledgements .....	vi
1 Chapter 1: Introduction .....	2
1.1 Background .....	2
1.2 Aims .....	2
1.2.1 Methodological aim .....	2
1.2.2 Staring and mutual gaze .....	3
1.2.3 Mutual gaze, task performance and social presence .....	4
1.2.4 Function of mutual gaze .....	5
1.3 Study outline .....	5
2 Chapter 2: Background and Literature Review .....	6
2.1 Social Presence .....	6
2.1.1 Why measure Social Presence? .....	9
2.1.2 How do you measure Social Presence? .....	11
2.2 Mutual Gaze .....	16
2.3 Mutual gaze and task performance .....	21
2.4 Gaze, task performance and social presence .....	24
2.5 From avatars to embodied conversational agents .....	28
2.6 Eye Tracking and ECAs .....	31
2.7 Avatars, agents and social interaction .....	34
2.8 Social facilitation and inhibition .....	38
2.9 Research questions .....	41

3	Chapter 3: Experiment 1 – Human-Human Study .....	43
3.1	Method.....	43
3.1.1	Independent Variables.....	43
3.1.2	Dependent Variables .....	44
3.1.3	Participants.....	44
3.1.4	Apparatus .....	44
3.1.5	Design .....	45
3.1.6	Stimuli.....	46
3.1.7	Procedure.....	49
3.2	Results .....	50
3.2.1	Hypothesis 1 – staring and mutual gaze.....	50
3.2.2	Hypothesis 2 – gaze and task performance .....	52
3.2.3	Further investigation into the dyads’ looking behaviour .....	57
3.2.4	Summary .....	60
3.3	Discussion .....	61
3.3.1	Staring and mutual gaze .....	61
3.3.2	Mutual gaze and task performance .....	62
3.3.3	Further investigation into the dyads’ looking behaviour .....	63
3.3.4	Next steps.....	63
4	Chapter 4: Experiment 2 – Human-Agent Experiment.....	64
4.1	Research questions – Gaze, task performance and social presence .....	65
4.2	Method.....	66
4.2.1	Independent Variables.....	66
4.2.2	Dependent Variables .....	69
4.2.3	Participants.....	71

4.2.4	Apparatus .....	71
4.2.5	Design .....	72
4.2.6	Stimuli.....	72
4.2.7	Procedure.....	74
4.3	Results .....	75
4.3.1	Task performance.....	76
4.3.2	Gaze.....	79
4.3.3	Task performance and Mutual gaze .....	86
4.3.4	Questionnaire .....	89
5	Chapter 5: Experimental Design .....	107
5.1	Exploratory Study.....	107
5.2	Design Requirements of the Human-Human Experiment.....	111
5.3	Additional requirements for the Human-Agent Experiment.....	113
5.4	Human-Human experiment: Independent variables .....	114
5.5	Building an experiment within Second Life.....	114
5.5.1	Tasks and Stimuli.....	117
5.5.2	Textures.....	118
5.5.3	Avatars .....	118
5.6	Experimental Design of Human-Agent Study.....	119
5.6.1	Audio instructions .....	119
5.6.2	Avatar controller .....	119
5.6.3	Questionnaire .....	120
5.7	Troubleshooting.....	122
6	Chapter 6: Discussion .....	123
6.1	Staring and mutual gaze .....	124

6.2	Mutual gaze, agency and looking condition: The 75% effect .....	125
6.3	The function of increased looking at the IG: Addressing a critique .....	128
6.4	Task performance and social facilitation.....	129
6.5	Social presence .....	133
7	Chapter 7: Conclusions and future work.....	137
7.1	Research questions revisited.....	137
7.2	Future work .....	138
8	References .....	145
9	Appendices .....	151
	Appendix A: Consent form for Experiment 1 .....	151
	Appendix B: Questionnaire for Experiment 1 .....	152
	Appendix C: Instructions for Agent condition (Experiment 2) .....	154
	Appendix D: Instructions for Avatar condition (Experiment 2) .....	154
	Appendix E: Questionnaire items (Experiment 2) .....	155
	Appendix F: Consent Form – Agent (Experiment 2).....	156
	Appendix G: Consent Form – Avatar (Experiment 2) .....	157
	Appendix F: Script 10c.50 (Experiment 2) .....	158

## Table of Figures

Figure 1: Instruction Giver's view of Instruction Follower's avatar.....	47
Figure 2: Instruction Follower's view of Instruction Giver's avatar.....	48
Figure 3: Mean % of trial IF spent looking at IG's avatar.....	51
Figure 4: Mean % of opportunities for mutual gaze taken up by the Instruction Follower .....	52
Figure 5: Areas of interest within the Instruction Follower's view .....	53
Figure 6: Mutual gaze and number correct in Staring condition .....	54
Figure 7: Mutual gaze and number correct in Not-Staring condition .....	55
Figure 8: Number correct and % of trial spent looking at task-irrelevant objects in Staring condition .....	56
Figure 9: Number correct and % of trial spent looking at task-irrelevant objects in Not-Staring condition.....	57
Figure 10: Mean % of trial spent in each looking behaviour in the Staring condition .....	59
Figure 11: Mean % of trial spent in each looking behaviour in the Not-Staring condition. This only includes times during which there were opportunities for mutual gaze. ....	60
Figure 12: Tracker's looking behaviour during experiment 2 .....	68
Figure 13: Participant's view of Tracker .....	73
Figure 14: Mean number of correct responses by Agency and Looking Condition ..	77
Figure 15: Mean RT since end of instructions by Agency and Looking Condition ..	78
Figure 16: Mean proportion of mutual gaze uptake by Agency and Looking Condition.....	79
Figure 17: Mean proportion of trial spent looking at each type of interest area in the Agent condition.....	80

Figure 18: Mean proportion of trial spent looking at each type of interest area in the Avatar condition.....	81
Figure 19: Mean % of trial spent looking at Stimulus Tiles by Looking Condition and Agency .....	82
Figure 20: Mean % of trial spent looking at Tracker by Looking Condition and Agency .....	83
Figure 21: Mean % of trial spent looking at Response Tiles by Looking Condition and Agency.....	84
Figure 22: Mean fixation duration by Interest Area type.....	85
Figure 23: Proportion of MG uptake and RT in 25% looking condition .....	88
Figure 24: Means and Standard Deviations of Factor Scores for each factor by Looking Condition and Agency .....	96
Figure 25: Distribution of Factor 3 scores by Looking Condition.....	97
Figure 26: Factor 3 Score and Reaction Time .....	99
Figure 27: Factor 3 Score and Reaction Time in 50% Looking Condition .....	100
Figure 28: Factor 3 Score and Reaction Time in 75% Looking Condition .....	101
Figure 29: Factor 1 score and Mutual Gaze Uptake in 75% Looking Condition ....	102
Figure 30: Factor 4 Score and Mutual Gaze Uptake in 50% Looking Condition....	103
Figure 31: Redundant condition: “From the wall behind me, select a domino with the same total number of dots as the one on the triangle”. .....	109
Figure 32: Non-redundant condition: “From the wall behind me, select a shape that you think best completes <i>this</i> sequence”.....	109
Figure 33: Experimental room .....	114
Figure 34: Stimulus and seating set-up .....	115
Figure 35: Experimental room set-up .....	116
Figure 36: A selection of stimulus tiles .....	117



## Tables

Table 1: Human-Agent experiment independent variables.....	66
Table 2: Mean and standard deviation for each questionnaire item by looking condition – Agent.....	89
Table 3: Mean and standard deviation for each questionnaire item by looking condition – Avatar.....	91
Table 4: Varimax rotated component loadings for 20 questionnaire items .....	94
Table 5: Paired samples t-tests for Looking Condition on Factor 3 factor scores .....	98
Table 6: Human-Human results .....	123
Table 7: Human-Agent behavioural results; X indicates no effect .....	123
Table 8: Human-Agent questionnaire results.....	124



# **1 Chapter 1: Introduction**

## **1.1 Background**

Recent technological advances have meant that face-to-face social interactions are increasingly being supplemented by computer- or video-mediated communication. Computer-mediated communication (CMC) and video-mediated communication (VMC), while practical (in terms of, among other things, financial and geographical considerations), may lack some of the non-verbal cues available in the real world, such as mutual gaze, or eye contact. These non-verbal cues are important for a successful interaction to occur. Social presence is a measure of how similar a mediated interaction is to face-to-face, in terms of perceptions of, and behaviour towards, an interlocutor. It would seem that, in order to maximise the potential of a mediated interaction, social presence should be as high as possible, and the measure of a successful mediated interaction would be one in which the interlocutors report a high level of social presence. Another measure of success during an interaction, specifically a task-oriented one, is task performance.

This thesis uses a paradigm within Second Life, an online virtual environment (VE) to examine the relationship between mutual gaze, task performance and social presence during computer-mediated interactions, with the aim of investigating the factors potentially underlying each of these variables.

## **1.2 Aims**

### **1.2.1 Methodological aim**

This research programme is composed of a methodological component, in service of a series of substantive aims. The methodological aim here was to design and implement a paradigm that would enable the investigation of users' eye movements within a virtual environment. The initial aim, therefore, was to identify areas in a previous paradigm (in terms of experimental design) that could, and needed to be, improved, in order to best investigate how users view a conversational partner and a virtual environment during a computer-mediated task-based interaction.

### **1.2.2 Staring and mutual gaze**

The first substantive aim was based on a gap in the research into mutual gaze. Mutual gaze can be defined as occurring “when two people are looking at each other’s eyes” (Bailenson et al., 2001; p.5). Much research exploring gaze – whether during video-mediated communication (VMC), computer-mediated communication (CMC), or face-to-face interaction – has assumed that an appropriate method for studying mutual gaze is to have one party in the interaction stare continuously at the other, the intention presumably being to maximise opportunities for mutual gaze (Argyle & Dean, 1965; Bailenson, Blascovich, Beall, & Loomis, 2001). Quite apart from the naturalness of constant staring by one conversational partner at another (i.e. does it replicate natural behaviour in real-life interactions?), could it be that it is actually discouraging mutual gaze; the participant is avoiding making eye contact because of the atypical human behaviour they are witnessing in their interlocutor (staring), since being the intense focus of someone else’s attention could be an uncomfortable experience? It could be that using this experimental method (staring) is, in fact, inadvertently impacting on the outcome measure (mutual gaze).

The first question, therefore, is: Is staring by one conversational partner at another a suitable method for investigating mutual gaze and, if not, how much looking by one conversational partner is appropriate for maximising mutual gaze between the dyad? If an instruction giver, for example, gazes continuously at an instruction follower, providing maximum opportunities, will this encourage maximum mutual gaze between the pair, and if not, then what is the optimum amount of time for one person to look at another during (in this context) a task-based interaction? In order to investigate this, the study used an avatar within Second Life who could be programmed to act as an agent (i.e. computer-controlled virtual human) to interact with a user within the environment. Overall, the first substantive aim was to establish that being continuously looked at by an avatar does indeed negatively impact on the amount of mutual gaze between a dyad, as well as establishing exactly what the optimum amount of looking by the avatar is for maximising mutual gaze.

### **1.2.3 Mutual gaze, task performance and social presence**

The next question expands on these results. Using the improved paradigm, the amount of instruction giver looking for maximising mutual gaze was investigated further, as well as the relationship between mutual gaze, task performance and perceived social presence. Social presence, sometimes known as co-presence, is a measure of how similar the experience of an interaction within a mediated environment is to a face-to-face interaction, the latter being the most socially present situation possible. If face-to-face interaction is the ideal, in terms of (among other things) access to non-verbal cues, maximising task performance and perceptions of an interlocutor, then it would follow that mediated interactions should aim to be as similar to face-to-face as possible.

Previous research has been varied, and sometimes inconsistent, with respect to defining, quantifying and measuring social presence. It has been suggested that existing self-report questionnaires do not capture the whole picture, and that behavioural measures (such as gaze or task performance) could detect aspects of social presence that are not identified by these subjective scales. The aim, therefore, was to find out what relationship, if any, there is between mutual gaze, perceived social presence and task performance. It was anticipated, based on previous research, that the degree of social presence that a user assigns to a conversational partner during a task-based computer mediated interaction may be related to the amount of eye contact that either party is willing to engage in with their interlocutor's avatar. A further possibility of a relationship between perceived social presence and task performance was also investigated. Any such relationships could present the possibility of using eye contact (perhaps in conjunction with task performance measures) as a behavioural measure of perceived social presence, which could supplement existing self-report questionnaires, thus providing a much-needed on-line method for measuring how socially present an individual perceives another being to be during mediated interaction.

In order to compare the efficacy of behavioural vs. self-report measures, actual social presence was manipulated by either telling users they were interacting with an avatar (i.e. human-controlled virtual interactant) or that it was an agent (i.e. computer

controlled virtual interactant). A questionnaire was chosen from existing literature and adapted and validated to ensure it was suitable for measuring social presence in the current paradigm.

#### **1.2.4 Function of mutual gaze**

The final question was raised by research suggesting that the function of gaze direction during social interaction can be divided into aspects related to social accessibility and those related to cognitive processes. If this is the case, then it follows that during a mediated interaction, the function of gaze direction can also be divided into these two categories. Using the data generated by the previous questions, the question of what, if anything, is driving mutual gaze during an interaction will be addressed.

### **1.3 Study outline**

The research is divided into two main studies. The first experiment, covered in chapter 3, was carried out on pairs of participants working together via avatars within Second Life to complete tasks. Their eye movements were recorded and were analysed post-interaction, along with their task performance. Chapter 4 reports on a study with a similar experimental set-up, although participants interacted with a computer rather than another human. Half of the participants were told that they were interacting with a human and the other half was told that it was a computer. Eye movements and task performance measures were analysed, along with responses to a social presence questionnaire. The research questions are outlined at the end of chapter 2. The next chapter reviews literature relevant to the research, as well as outlining the motivation behind the studies.

## **2 Chapter 2: Background and Literature Review**

There is an enormous amount of research relating to this thesis, with many interesting issues. For our current purposes, however, the most relevant points of intersection between them are discussed. The main areas to be covered in this chapter are: a review of recent literature relevant to the study of social presence – including why it is of interest and how it is measured; a review of studies into mutual gaze – including recent and classic research; an outline of work on embodied conversational agents (ECAs) and human-controlled avatars; how eye tracking measures have been applied to ECA research; and finally how avatars and agents have been used to investigate social interaction, with particular emphasis on social facilitation.

### **2.1 Social Presence**

Social presence (also sometimes known as co-presence) is an area of growing interest, with many definitions and measures having been postulated in recent years. Despite this, there has been a lack of any overarching measurement, or even conclusive definition. For example, the terms ‘Social presence’ and ‘copresence’ are used interchangeably by some researchers, whereas others maintain that social presence is made up of several dimensions, one of which is co-presence – the feeling that you are not alone (Harms & Biocca, 2004). Social presence has been defined as “the degree of salience of the other person in the interaction” (Short, Williams, & Christie, 1976), the feeling, when interacting with an interlocutor, of “the degree of initial awareness, allocated attention, the capacity for both content and affective comprehension, and the capacity for both affective and behavioral interdependence with said entity.” (Harms & Biocca, 2004) and the “sense of being together” (de Greef & IJsselsteijn, 2000). Biocca et al. (2003) separate definitions of social presence into three categories:

1. Those related to copresence (colocation, mutual awareness); e.g. “the feeling that the people with whom one is collaborating are in the same room” (Mason, 1994) or “the degree to which a person is perceived as a ‘real person’ in mediated communication” (Gunawardena, 1995). De Greef and IJsselsteijn’s “sense of being together” also fits into this category.

2. Those related to psychological involvement:

- Perceived access to another's intelligence, e.g. "...the degree to which the user feels access to the intelligence, intentions, and sensory impressions of another." (Biocca, 1997)
- Short et al.'s "salience of the other person" (see above) and
- *Intimacy*: the ability of a medium to convey physical distance, eye contact, smiling, based on Argyle's (1965) theory of eye contact, distance and affiliation
- *Immediacy*: Wiener and Mehrabian's concept enhances social presence, and can be defined as a "measure of the psychological distance that a communicator puts between himself or herself and the object of his/her communication" (Gunawardena & Zittle, 1997), i.e. the "psychological distance" between interactants (Wiener & Mehrabian, 1968).
- Mutual understanding: "... the ability to make one's self known under conditions of low media richness" (Savicki & Kelley, 2000).

3. Those related to behavioural engagement:

- To experience social presence is to be able to "effectively negotiate a relationship through an interdependent, multichannel exchange of behaviours" (Palmer, 1995).

It may be appropriate to include the possibility of cognitive measures, such as task performance, although there does not seem to be much precedent for their use within social presence research.

As has been demonstrated, there are several facets to social presence, perhaps due to the wide ranging scope of mediated social interactions within which it is measured, thus explaining the lack of overarching measurement or conclusive definition. The definition of social presence may well be specific to the medium being used to experience an interaction. In fact, early definitions of social presence maintained that it was a function of communication medium alone, rather than the content of the interaction (Short et al., 1976). It was thought that the lack of access to non-verbal cues in early CMC would result in a paucity of social information, which would



consequently determine how individuals interact and automatically lead to decreased social perceptions. There is evidence to suggest, however, that affect during CMC can be increased without non-verbal cues. In spoken CMC, for example, participants may increase their use of socio-emotional cues verbal cues to compensate for a lack of non-verbal ones (Hiltz, 1994). In text-based CMC, affect can be increased by the use of emoticons or emojis (Gunawardena, 1995); this is something that has become invaluable with the more recent widespread use of text-based social interaction (social media, email, text messaging). Reid (1994) observes that users of the early text-based virtual worlds, Multi-User Dimensions (MUDs), developed a shared language and textuality which, despite lack of access to non-verbal cues, allowing successful social interaction and development of online communities. Gunawardena (1995) examined whether social presence was a product of the communication medium, or users' perceptions of that medium. She carried out two studies, in which students responded to 17 5-point bipolar scales about their "current feelings" towards the medium of CMC during computer-mediated conferences (p. 150). The responses, therefore, related to the students' use of CMC both during and prior to the conferences. The bipolar scales included items such as sensitive-insensitive, sociable-unsociable and helpful-hindering. Study 1 looked at students after one computer conference in 1993. Study 2 compared two conferences: one involved online group discussion related to distance education and the second was a collaborative learning exercises. The results of the studies suggested that, despite the lack of non-verbal and social context cues, participants were able to "create social presence by projecting their identities" (p. 163). The evidence points to the fact that, contrary to Short et al.'s idea that the type of media is wholly responsible for determining social presence, the contents of a social interaction are also, in fact, important.

Another determinant of social presence may be the purpose of an interaction. Task-based communication, for example, may focus on behavioural aspects of social presence, whereas a purely social interaction may concentrate more on psychological involvement. During this research, social presence (rather than co-presence) will be referred to and investigated. For the purposes of this thesis, we have defined social presence as a measure of how similar the experience of a computer-mediated social

interaction is to that of a face-to-face interaction (the most socially present situation) in terms of perceptions of and behaviour towards an interlocutor. This definition relates to the purpose of the interactions and what we intend to measure; participants will be completing tasks and reporting perceptions of their interlocutor, therefore our definition includes ‘perceptions of and behaviour towards an interlocutor’ to encompass the comparison of behavioural and self-report measures within our study.

### **2.1.1 Why measure Social Presence?**

Since the early 1990s, the ubiquity of the Internet has resulted in an increase in use of online interactions, particularly via virtual environments (VEs). A VE can be defined as “synthetic sensory information that leads to perceptions of environments and their contents as if they were not synthetic” (Blascovich et al., 2002). An immersive virtual environment (IVE) is a VE in which the user has the perception of being physically present within the space and able to interact with the environment. A shared VE is one in which more than one user is present within the virtual space, each often appearing as avatars, or virtual beings. Virtual environments have moved from early Multi User Dungeons or text-based multi-user virtual worlds to current 3D virtual environments used for socialising, learning and business.

Second Life (SL) is a “3-D virtual world entirely built and owned by its residents”, where users are able to interact with other users and agents within a 3D virtual environment (see <http://secondlife.com/>). Interaction occurs through a user’s avatar, whose appearance can be personalised to the user’s taste, from altering the height, clothes or hair colour of the avatar to giving it wings, or even the head of an animal. Using our definition of social presence as the extent to which one person’s experience of another in a computer mediated interaction is similar to being in a face-to-face interaction, it starts to become clear why social presence should be quantified and measured. Computer- and video-mediated interactions are becoming increasingly necessary, due to, among other things, geographical disparity between interlocutors (leading to financial and temporal issues). If the ideal is a face-to-face interaction then it is important to establish what needs to be improved in order to make a mediated interaction as similar to face-to-face as possible, and consequently as fruitful as it can be. If this can be quantified by a measure of social presence, then

the quality of an interaction can, in turn, be quantified, thus enabling communication researchers to design and implement the most effective mediated communications possible (i.e. the most similar to face-to-face, in terms of perceptions of and behaviours towards an interlocutor).

Bailenson and colleagues report that:

*‘Co-presence, also called social presence, occurs when a person uses a virtual environment that contains one or more virtual agents and behaves as if he or she were interacting with other veritable human beings. In other words, when a virtual agent influences a user of an IVE in a similar manner than a real person would, that user experiences high co-presence.’*

(Bailenson, Beall, & Blascovich, 2002)

The authors of the above paper recommend a combination of behavioural and self-report measures of social presence, which will be discussed in due course. Nowak argues that measuring an individual’s sense of social presence during interaction with an interlocutor is one way to determine the efficacy of a communication medium, in terms of ability to fulfil communication goals (Nowak, 2001, May). According to this definition, therefore, measuring presence (copresence, social presence and presence as transportation in this case) can provide a measure of the usability, and an overall evaluation, of a communication interface. According to some definitions, presence does not occur when a user is aware that the experience is mediated (Lombard, 2000).

We have defined social presence as a measure of how similar the experience of a mediated interaction with an avatar or agent is to that during a face-to-face interaction (the most socially present situation) in terms of perceptions of and behaviour towards an interlocutor. A high level of social presence during a mediated communication would elicit similar perceptions and behaviours as a face-to-face interaction would. In physically sharing a space with someone, and being able to hear, see and touch them and interact in a believable way, a high level of social presence can be achieved. As these non-verbal communication channels become unavailable (due to differing levels of mediation), the perception of social presence is reduced. Thus, by ensuring that some of these cues are available in a mediated

interaction, levels of social presence can be increased. In quantifying and measuring levels of social presence, we have an idea of how similar an interaction is to the ideal of face-to-face.

### **2.1.2 How do you measure Social Presence?**

As previously mentioned, there are several different researchers who have aimed to quantify and measure social presence, and within much of this research there is disagreement as to the best way to measure it. Methods fall into one of two broad categories: *subjective self-report* and *behavioural* measures.

#### **2.1.2.1 Subjective measures**

Subjective measures typically consist of self-report questionnaires, using a Likert-type scale. These tend to be presented post-interaction, due to their reflective nature. Biocca et al. (2003) summarise different types of subjective self-report tools, and the dimensions which they measure, as follows:

1. Perceived social richness of the medium:
  - “Media having a high degree of social presence are judged as being warm, personal, sensitive and sociable” (Short et al., 1976)
2. Involvement, immediacy or intimacy:
  - Typically involving a five-point Likert scale, assessing a conversational partner’s involvement, interest or level of emotion relating to the conversation; based on the original relational communication schemas devised by Burgoon & Hale (1987).
  - For example, measuring the intimacy of the medium involves quantifying its capacity to convey non-verbal information, which contributes to perceptions of its social presence (Gunawardena & Zittle, 1997).
3. Social judgements of others:
  - *Homophily* is a measure which “attempts to capture the sense of feeling similar to the other in attitudes, behaviors, or emotions” (Biocca et al., 2003).
  - Uses scales adapted from McCroskey et al. (1975).

One of the main issues with self-report measures is that they are administered post-interaction, and therefore can only identify a user's recollection of social presence perceptions, rather than tracking them during the communication. This could lead to some elements being forgotten or overlooked, suggesting that additional methods should be explored.

#### **2.1.2.2 Self-report, behavioural and cognitive measures**

Historically, there has been less focus on behavioural measures of social presence than subjective, self-report. This may be in some part due to the availability of technology for capturing behaviour (for example, eye trackers and virtual environments) but more recently, that focus has shifted to some extent. Bailenson and colleagues made a direct comparison between subjective (self-report) and behavioural measures of social presence (Bailenson et al., 2004). They measured the interpersonal distance (IPD) between a participant and either an embodied tutoring agent or an unfamiliar embodied agent while the user was walking through an IVE. They found that more space was given to the embodied tutor than the unfamiliar agent. Self-report measures, however, revealed no difference between the subjective reactions to the tutor and the unfamiliar agent. The authors suggest that “nonverbal behaviour may be a more sensitive measure of the copresence and general influence of embodied agents than self-report measures” (Bailenson et al, 2004; p.216). It may be, however, that there were too many variables within the study, and that separating and varying each one systematically would produce a clearer picture of what can be used to accurately measure presence. The authors recommend that the use of behavioural, as well as other measures would be advantageous in the study of social presence.

In another study, Bailenson and colleagues manipulated the mutual gaze of virtual humans as participants interacted with them within an immersive virtual environment (IVE). Here, again, the assumption is that constant looking by a virtual human somehow equates to mutual gaze (Bailenson, Blascovich, Beall, & Loomis, 2003). In the first study, participants moved around a virtual human in a virtual room, and in the second study, the virtual human traversed the virtual room as well. The gaze behaviour of the virtual human was varied, as well as whether it was allegedly

controlled by a human (an avatar) or a computer (an agent). There were 2 gaze conditions. In the high gaze condition, where the participant was “engaged in mutual gaze”, the virtual human “blinked his or her eyes at a natural rate and turned his or her head to gaze at participants’ faces as they traversed the IVE” (i.e. stared at the participant; Bailenson et al., 2003; pp.819&822, respectively). In the low-level gaze condition, the virtual human had its eyes closed, and the head did not turn. These seem to be extreme gaze conditions, neither of which approximate particularly well to natural human behaviour, and this supports the idea that an optimum amount of gaze by an agent/avatar, or at least a typical amount, still needs to be established, in order to maximise mutual gaze. In the first experiment it was found that more space was given to the front of virtual humans than the back, and that “participants gave more personal space to virtual agents who engaged them in mutual gaze” than to agents who did not (Bailenson et al., 2003; p.819). They did not find a difference between the gaze conditions when participants thought they were interacting with avatars. This means that when participants believed they were interacting with a person, they maintained appropriate interpersonal distance, whatever the gaze behaviour. When interacting with a computer (agent), however, it needed to have high behavioural realism in order for the appropriate interpersonal distance to be maintained. Here, it is assumed that the looking in the mutual gaze condition displays high behavioural realism, and in the low gaze condition demonstrates low behavioural realism. Indeed, they found that social presence ratings were higher in the high gaze condition than the low gaze condition. This is somewhat surprising; we have already touched on the idea that staring is likely *not* to equate to typical looking behaviour, and therefore high behavioural realism. This idea will be extended further in the Staring and mutual gaze section.

These results show evidence for the threshold model of social influence, in that the correct combination of social presence and behavioural realism is required in order for an agent or avatar to wield social influence over an interlocutor. This theory will be discussed later in the literature review (see section 2.8: Social facilitation and Inhibition). It is of note to question here whether it is actually possible for one person to *engage another in mutual gaze* due to the implicitly mutual nature of mutual gaze. They found no relationship between distance and social presence ratings. In the

second experiment, they found that participants were more likely to show avoidance behaviour to the computer-controlled agents than what they thought were human-controlled avatars. They suggest that this could be because avatars would not have been expected to walk into someone (since it is not something that one would expect of a human) but that agents might. Again the authors conclude that behavioural measures may be more effective in establishing levels of social presence than questionnaires alone. In terms of an experimental design, manipulating the level of social influence that the virtual human has over the user, whilst keeping all other sensory information constant (i.e. saying that it is an avatar vs. an agent whilst keeping the rest of the design constant) would seem to be an ideal way of manipulating how socially present the virtual human actually is. From this, it should be possible to establish the validity of self-report measures of social presence.

In a further paper, Bailenson and colleagues discussed the division of co-presence research in virtual environments into 2 areas – self-reported *perceptions of* an agent (the feeling that you are not alone) and social, task-related and physiological *responses to* that agent (measured by physiological data, verbal/non-verbal behaviours, task performance, cooperation, liking or other sociological responses) (Bailenson et al., 2005). They aimed to compare these two types of co-presence measures, looking at perceptions via self-report measures, and also at cognitive measures as well as behavioural responses to the agent. The self-report measures (perceptions) consisted of how much an individual felt as if there was another social entity in the virtual environment, liking of an agent and how willing you would be to perform embarrassing behaviours in front of it. It was hypothesised that co-presence and liking would be positively correlated, and that copresence and willingness to perform embarrassing behaviours would be negatively correlated. The behavioural measure was interpersonal distance. It was hypothesised that co-presence would be positively related to hesitation in approaching an agent, interpersonal distance and maintenance of IPD. The researchers also looked at cognitive markers – would co-presence increase or decrease memory for information relating to the agent? It was found that behavioural realism had less of an influence on perceived copresence when the agents looked human-like than it did when the agents looked non-humanlike; ratings of co-presence were lowest when there was a large mismatch

between the appearance and behavioural realism of the agent. Participants were less willing to perform embarrassing acts in front of agents that they had rated as having a high level of co-presence, and rated them as more likeable than agents that they rated as having low co-presence.

In terms of task performance, it was found that during interaction with agents that were rated as having a higher level of co-presence, there was some evidence of participants scoring better on a memory test. This correlation was not, however, statistically significant. The first behavioural measurement of interpersonal distance (minimum distance between user and agent) was not affected by behavioural realism, and there was no relationship between ratings of co-presence and interpersonal distance (minimum distance and number of returns to the agent). The user showed more hesitation in approaching the agent (i.e. number of returns to the agent) whose behaviour was more humanlike (compared with random behaviour). This is a complex study, and the authors themselves conclude that many variables from different channels can make interpretation of the results more complicated. It does, however, provide interesting results which could be extended individually within the paradigm. Further investigation into the task-performance-co-presence ratings results, for example, could assist in the overall understanding of factors which affect, or are affected by, social presence.

Other, more recent, studies have also investigated the utility of behavioural and subjective measures of presence. One such study investigated the relationship between, among other things, presence – social presence and temporal presence (“the feeling of how much the player felt present in the time of the game”) and behavioural measures (Von Der Pütten et al., 2012). The behavioural measures used were pointing and verbal responses towards virtual content within an augmented reality scene. Although a significant negative correlation was found between the behavioural responses and perceptions of temporal presence, no relationship was found between behavioural and social presence measures. They suggest, however, that behavioural measures do have the potential to be used to predict feelings of presence within augmented reality games.



In terms of an overarching method for measuring social presence, it is possible that the existing self-report measures fail to fully capture all of the facets of social presence. There are several questionnaires designed to measure the dimensions of social presence (De Kort, IJsselstein, & Poels, 2007; Harms & Biocca, 2004; Short et al., 1976; Witmer & Singer, 1998). Of these, Harms and Biocca aimed to develop and test the validity of a measure of social presence dimensions, including co-presence (the feeling you are not alone), attentional allocation, perceived message understanding (how well you understand your conversational partner's message and vice-versa), perceived affective understanding (how well you understand your partner's emotional and attitudinal state and vice-versa), perceived affective interdependence (how much one interactant's emotional and attitudinal states are affected by the other's and vice-versa) and perceived behavioural interdependence (how much you believe your behaviours are affected by the other and vice-versa). To test the 80 items that were created to reflect the identified dimensions, the scale was validated on face-to-face, text-mediated and video-mediated interactions. From this, the authors concluded that their scale was a valid measure of the various dimensions of social presence. If, as previously discussed, an ideal way to measure social presence is with a combination of self-report and behavioural, as well as the possible inclusion of cognitive measures, then this scale may not tell the whole story. It appears, therefore, that an initial validation of a self-report measure of social presence should be compared with an additional, behavioural measure.

## **2.2 Mutual Gaze**

Non-verbal communication is an important contributor to successful social interaction. Gaze direction, in particular, provides rich social information, such as social accessibility: *mutual gaze*, or eye contact, can indicate that a conversational channel is open, and that an interlocutor is willing to engage, or continue to engage, in an interaction. Although they have slightly different implications, the terms 'mutual gaze' and 'eye contact' will be used interchangeably throughout this thesis. *Joint attention*, or being aware of a conversational partner's eye movements, and

consequently focusing on the object of their attention, is a skill that is developed in infancy (Corkum & Moore, 1998), and is widely used during conversation. For example, we can infer the object a conversational partner is referring to by following their gaze. Mutual gaze has also been reported to facilitate performance on certain cognitive tasks. Fry and Smith (1975) for example, found that increased eye contact resulted in better task performance on a digit encoding task.

Kendon (1967) reported some of the functions of gaze direction during social interaction, dividing them into aspects related to social accessibility – an indication of a willingness to engage in, and commitment to continue, an interaction – and functional purposes – changing gaze direction to signal to another person what you are referring to. Kendon further reports that eye contact between a dyad is also regulated by other social aspects, such as how much one person is drawn to another in an affiliative way (higher affiliation will lead to more EC), or to what extent one is in competition with another (more competition will lead to more EC). When observing 7 dyads during videoed conversations, they found that there was a large variation in the amount of time spent looking at the other, the lowest being 28% and the highest over 70%. He also reported that each instance of mutual gaze was short – no more than a second at a time – and that duration of time spent looking at another participant was related to the amount of time the other looked at them. In terms of listeners, Argyle and Dean (1965) found that they spent on average 75% of a dyadic conversation looking at a speaker. Nielsen (1963) reports that people spend less time looking at their conversational partner when speaking than when listening.

Research into mutual gaze can be divided into 3 types: that in relation to face-to-face communication, video-mediated communication (VMC) and computer-mediated communication (CMC). Within these areas, the definitions and measurements of mutual gaze have been diverse. Argyle and Dean (1965) aimed to outline the empirical determinants, psychological processes and functions of eye contact (mutual gaze) in face-to-face interactions. They report on some of the functions of eye contact, including:

- Information seeking, including feedback at the end of sentences in order to judge responses of others;

- Signalling that the conversational channel is open, informing that the conversation can proceed (if one party looks at a 3rd person, it may indicate that the channel is closed);
- Information about the focus of the other's attention;
- Concealment and exhibition – eye contact as proof that you are being seen, or, conversely, less eye contact if you do not want to be seen (here, they draw on research that showed that subjects who had been induced to cheat looked less at their conversational partner's eyes) and
- Establishment and recognition of social relationships.

The authors discuss the affiliative conflict theory, which posits a tension between approach and avoidance. According to this theory, there is a need for feedback in an interaction, as well as affiliative (social) needs, leading to approach. Eye contact “can be used as a reinforcer in the operative conditioning of verbal behavior” (Argyle & Dean, 1965; p. 292). A fear of being seen and revealing inner states lead to rejecting, or avoidance, responses. If this theory is true, then there should be an equilibrium level of eye contact for one person in social contact with another, which will be mediated by other behaviours linked to affiliative motivation. “Thus there will be an equilibrium point of physical closeness, of intimacy of conversation, and amount of smiling. The more these behaviors occur, the more affiliative motivation is satisfied, but if they go too far, anxiety is created” (Argyle & Dean, 1965; p.293). According to this equilibrium equation, if one of the components is changed (such as physical proximity), then another component in the equation will need to alter to compensate, thus maintaining equilibrium.

To test the hypothesis that increasing spatial proximity between two individuals decreases eye contact, they observed 2 people in conversation, one of whom was a confederate who gazed continuously at the other conversational partner (the subject), thus ensuring that the amount of eye contact was entirely under the subject's control. They varied the distance between the two seated individuals – 2, 6 and 10 feet apart – and their chairs were placed facing each other in different orders for different pairs. In the final version of the experiment, the chairs were placed at 90° and behind tables to make the eye contact more ‘voluntary’ (i.e. to make it easier for participants to

avert their gaze) and the continuous gazing of confederate less obvious. They measured the amount of eye contact during a 3-minute conversation, as judged by an observer at a distance. They found that reducing eye contact makes greater proximity possible and that greater proximity reduces eye contact between two individuals, supporting the affiliative conflict theory.

Critics of this study suggested that rather than supporting the model of equilibrium for intimacy, “these results may simply demonstrate that, with increasing distance, observers find it more difficult to distinguish between what is eye-contact and what is merely gaze in another direction, and increasingly record as eye-contact much that is not” (Stephenson & Rutter, 1970). They suggest that the results may have been an artefact of observer performance, rather than subject performance; repeating the experiment in order to find out if the observer could judge the difference between eye contact and looking at a conversational partner’s ear or shoulder (for example). In response, Argyle objected to this interpretation of his results on the basis that the assumption that “during social interaction real people spend a significant proportion of the time looking at each other’s ears, shoulders or adjacent areas” is false. He maintained that Stephenson & Rutter’s critique “would only be applicable in a world in which ear-glances and shoulder-glances are common”(Argyle, 1970). This is an interesting issue, which highlights the obvious benefits of the advances in eye tracking technologies, in that eye movements can now be measured accurately with greatly reduced potential for error. Whatever the truth in this matter, the questioning of the validity of measures of mutual gaze during face-to-face interaction show the importance of developing a paradigm that can conclusively determine whether, at any one time, an individual is making eye contact with a conversational partner.

Monk and Gale define mutual gaze as “knowing whether someone is looking at you” (Monk & Gale, 2002). They looked at the benefits of having access to mutual gaze and full gaze awareness, compared to audio alone during VMC. Full gaze awareness was defined as the “ability to gauge the current object of someone else’s visual attention” (Monk & Gale; p.2). They found that, in terms of the efficiency of the conversation, full gaze awareness reduced the need for the use of verbal language, and that it was far superior in this respect to mutual gaze alone. One of the problems

in this study was that VMC does not usually support mutual gaze; the discrepancy between the position of the camera and the image of the other person's eyes means that it is not possible to simultaneously look at a conversational partner's eyes as well as the camera. Access to mutual gaze during VMC could only be achieved by using a video tunnel set-up (a specific method that enables eye contact during VMC). Due to the previously discussed importance of the functions, and social benefits, of mutual gaze during an interaction, it appears that it, as well as full gaze awareness should be included in an interface to maximise access to nonverbal cues during an interaction, although the authors were unable to report on the benefits of having access to mutual gaze and full gaze awareness at the same time since accessing mutual gaze was not consistent with full gaze awareness in this set-up. This presents a problem since it would, in Monk and Gale's paradigm, involve extra windows to be open to the viewer, meaning crowded screen space and a requirement for attention shifts between the windows. This study highlights the importance of developing a platform that supports mutual gaze, as well as full gaze awareness.

Yee and colleagues investigated the persistence of nonverbal social norms in online environments (Yee, Bailenson, Urbanek, Chang, & Merget, 2007). They hypothesised that, in line with Argyle's equilibrium theory, the closer 2 avatars are within an online environment, the less likely they are to maintain eye contact. To establish this, researchers collected observational data from snapshots of dyads within Second Life, recording the direction that the avatars were facing and the distance between them. One problem with this study, however, is that the avatars' behaviour is not a direct indication of the behaviour of the user, since much of the avatar behaviour (including some aspects of eye/head movements) is automated by Second Life and not controlled by the user, although head movement can be controlled by the user. Another issue with this study is that other, social aspects, including familiarity between the two individuals, have also been shown to regulate eye contact and interpersonal distance (Argyle & Dean, 1965). This could not be determined by simply observing dyads. Although this is an interesting study, it highlights the importance of being able to measure where the user is looking during an interaction, as opposed to their avatar.

Bailenson and colleagues investigated how much space in an immersive virtual environment (IVE) participants left around an agent that constantly maintained eye contact with them, “that is, mutual gaze behaviour”, compared with an agent that had its eyes closed (Bailenson, Blascovich, Beall, & Loomis, 2001). They varied the photographic realism of the agent, as well as its gaze behaviour. In this investigation into mutual gaze (as well as in Argyle and Dean’s study), the paradigm involves one member of a dyad staring continuously at the other, and any negative consequence of this looking behaviour on mutual gaze is not addressed. There is no discussion of the potential impact of staring on mutual gaze, and if this may have produced experimental artefacts. Argyle and Dean’s study at least employed some measure of how much actual mutual gaze there was, although Bailenson did not. Quite apart from the question of how staring impacts on amount of mutual gaze, there is the question of how natural a behaviour (i.e. how similar to typical human behaviour) constant gaze from an agent is, and how this pattern of looking affects perceptions of an agent, as well as performance in a task that is being carried out. Kendon suggests that “To be subjected to the continual gaze of another is a very unnerving experience, for to be the object of another’s attention is to be vulnerable to him” (Kendon, 1967). In terms of perceptions of an agent, Bailenson and colleagues found that perceptions of co-presence were lowest when there was a large mismatch between the appearance and behavioural realism of an agent (Bailenson et al., 2005). If constant looking by an agent is not behaviourally realistic (typical of natural human behaviour), then how will this affect the perception of it? It would therefore be of interest to discover the effect of constant gaze by an agent on the amount of mutual gaze between the dyad, as well as how this constant gaze affects the perception of an agent.

### ***2.3 Mutual gaze and task performance***

As previously discussed, mutual gaze has been known to serve several functions, one of which is, under certain circumstances, the facilitation of task performance. This facilitation has been found in various areas of eye movement research. In one study, researchers investigated how looking at a camera during a presentation over a video link affects information recall (Fullwood & Doherty-Sneddon, 2006). They report

that in face-to-face interaction, mutual gaze can facilitate encoding and subsequent recall of material. They suggest that this is because gaze functions as an arousal stimulus, thus increasing focus of attention and subsequent recall of material. They wanted to find out if gaze at a camera during VMC would result in the same benefits. During a presentation a confederate gazed at the camera either during 30% of their presentation to a subject, or not at all. Here, the authors maintain that gaze into the camera gives the impression of gazing in the user's direction. They found that more gaze by the confederate into the camera resulted in higher recall of the material by the subject. They also compared the results from the same presentations with the video switched off, to act as controls, and found no difference between the two conditions, meaning that any differences between them had not arisen as a result of differences in audio signals. They conclude that "the perception of gaze aversion over a video link (a consequence of the salesman not looking into the camera) has a negative impact on information recall." (Fullwood & Doherty-Sneddon, 2006; p.167)

This is an interesting result, as it implies that access to eye contact in VMC (as has been shown in face-to-face interactions, and therefore potentially CMC) could affect task performance. If access to mutual gaze can, indeed, facilitate task performance, it would be pertinent to find out how to maximise the amount of mutual gaze between a conversational pair (dyad). In this study (Fullwood & Doherty-Sneddon, 2006), the gaze by the confederate is at a more behaviourally realistic level than in Bailenson's study, although there is still no measure of how much actual eye contact there was between the individuals, only assuming that, since the participant is watching the presentation, mutual gaze is achieved during the confederate's looking. In the previously mentioned Fry and Smith's (1975) study into the facilitation of task performance by eye contact, they merely state that "Eye contact was manipulated" during the experiment, with an instruction giver giving her conversational partner either "as much eye contact as possible" or "as little eye contact as possible", depending on the condition (p. 2). One of the fundamental aspects of mutual gaze is that it is a joint action – one cannot independently engage in mutual gaze, and therefore cannot give (or be given) eye contact, as it is an inherently mutual activity. As with Argyle and Dean (1965) and Bailenson et al.(2001), the assumption is that a single person has the ability to independently control the amount of mutual gaze that

occurs between himself and a conversational partner. It assumes that all one must do to in order to engage in mutual gaze is to stare at a conversational partner.

In another paradigm, Vuilleumier et al. (2005) investigated the effects of gaze perception on face processing. Faces were presented, in the form of photographs, with either eyes towards a viewer, or averted, and either a full-frontal or a  $\frac{3}{4}$  view of the faces. They measured online gender judgement (i.e. during the viewing) and post-test, surprise facial recognition tests (i.e. without warning or preparation). They found better gender judgement when the faces were presented with eyes towards the viewer than when they were averted, and this effect was particularly strong when the face being viewed was of the opposite gender to the observer. Likewise, post-viewing recognition of the faces was higher when eyes were presented towards the viewer, especially, again, when the face being presented was of the opposite gender to the viewer. This is further evidence, within a different medium, for the facilitation of task performance by access to eye contact. It would be interesting, again, to see if this facilitation would occur in other areas of research, such as during an interactive task-focused CMC.

Schneider and Pea (2013) investigated how access to the gaze of an interlocutor affects performance on collaborative tasks. Remote dyads worked together to study diagrams, with the aim of collaboratively explaining how visual information is processed by the human brain. Half of the dyads had access to their interlocutor's gaze (joint attention), and the other half did not. They found that those dyads who had access to their interlocutor's gaze had more effective collaborations and showed increased learning, compared to those without. Bailenson and colleagues provide further evidence of a more efficient interaction when a user has access to non-verbal cues (Bailenson et al., 2002). They used avatars within an immersive virtual environment (IVE) to investigate the effects of three different types of avatar head-movement (rather than eye movement) on perceptions of and behaviour towards an interlocutor. Three remote participants met within the virtual environment; participants were represented by human-like avatars with either head movements rendered in real time, a human-like avatar without head movements or a voice only. They found that the moving head avatars elicited higher co-presence ratings and



liking, as well as attracting more looking towards the heads. A lower percentage of the time was spent speaking during interactions with moving heads compared with the other conditions. Although this study uses head movement, rather than eye movement, it adds weight to the idea that there could be some relationship between non-verbal communication and task performance.

What can be seen is that there are several research areas in which task performance on various types of task has been facilitated by access to mutual gaze. There is even evidence from human-robot interaction that increased gaze by a robot towards a listener during story telling can increase post-story recall (Mutlu, Forlizzi, & Hodgins, 2006). Other research, however, has found mutual gaze to be detrimental to task performance, especially during cognitively demanding tasks (Buchanan et al., 2014). It appears that different patterns may emerge in different research areas, and depending on the task involved. In order to maximise the efficiency of any given task-based interaction, it is important to discover whether performance can, indeed, be facilitated by mutual gaze since it cannot be taken for granted that this pattern carries across paradigms, types of communication and research areas.

## **2.4 Gaze, task performance and social presence**

A further study looked how rapport relates to mutual attentiveness, positivity and coordination when interacting via avatars in virtual worlds (Wang & Gratch, 2010). They define rapport as a “subjective feeling of connectedness”, which arises in face-to-face interactions from the non-verbal behaviours: mutual attention, positivity (smiling and head nodding) and coordination (behaviours that demonstrate the couple are functioning in a coordinated way, such as postural mirroring and interactional synchrony (Wang & Gratch, 2010)). They measured rapport using a 10-item rapport scale, including items such as *I think the listener and I established rapport* and *I felt I was able to engage the listener with my story*. This measure would appear to equate to some degree with the concept, or some of the facets, of social presence. Participants told a story to what they believed to be another participant’s avatar, but which was actually a computer-controlled agent. The agent systematically varied its levels of attentiveness, positivity and coordination. There were 3 levels of agent gaze behaviour – responsive, staring and ignoring. In the

responsive condition, the agent stares continuously at the speaker and “exhibits attentive listening behaviors including head nods and posture mimicking that have previously been demonstrated to create self-reported feelings of rapport”, as well as idle-time behaviours (Wang & Gratch, 2010). In the staring condition, the agent’s behaviour is the same as in the responsive condition, but without the listening behaviours; in the ignoring condition he looks around the room, only looking randomly at the speaker.

The researchers examined the participants’ visual attention (by annotating a video of movement and position of the participant’s head during the interaction) and found no effect of agent gaze condition on how many times the participant gazed at the agent. There was, however, a trend towards more looking at the agent in the responsive condition than in the other two. The proportion of interaction during which the participant gazed at the agent was very similar in the staring and ignoring conditions. This would suggest that the addition of attentive listening behaviours could increase the amount of attention one is willing to afford a conversational partner. It also suggests that staring at someone without attentive behaviours is as effective as ignoring them in averting a conversational partner’s gaze. They report on one of their earlier studies with a similar experimental set-up that found the average amount of time that a human speaker spent looking at a human listener was 65%. They found no significant difference in overall gaze duration between the agent looking conditions, but gaze duration over time was different between the three conditions. They state that the gaze duration in the responsive condition remains high throughout and in the ignoring condition remains low throughout. The gaze duration in the staring condition is high at the start (perhaps to try and ascertain whether the partner is listening, since there are no non-verbal behaviours showing attentiveness) and falls to the level of the ignoring condition throughout the trials (although they don’t report post hoc tests so it’s difficult to draw any statistically significant conclusions). In terms of feelings of rapport, they found significantly higher levels in the responsive condition than in either the staring or ignoring conditions, but no difference between the latter two. Furthermore, they found the agent in both the staring and ignoring conditions to be more distracting than in the responsive condition.

As previously described, Wang & Gratch's study investigates the looking behaviour of a speaker during interaction with an agent displaying varying amounts of attentiveness. The differences in amount of looking at listener between conditions were not significant, only showing trends, apart from the gaze duration, although no post hoc tests are reported. Overall, it is a useful study for helping inform the design of gaze behaviour for computer- and human-controlled agents in order to maximise attention and feelings of rapport. The next logical step would seem to be to record eye movements with an eye tracker during interaction with the agent, since it would provide a more accurate record of gaze direction / duration than movement and position of the head. Further investigation may reveal more about the relationship between gaze and subjective perceptions of rapport, for example, where a person is looking while being stared at, and how this relates to perceptions of rapport.

Many interactions with agents and avatars involve a task (completing a purchase, playing a game, solving a query). It would be interesting to find out how the agent behaviour impacts on task performance, since the task in the current study was merely to recount a story to the agent. It would also be interesting to find out what the effect of speaker behaviour is on the listener's social perception and attention, as well as task performance, since interactions with agents and avatars do involve them speaking as well as listening.

As previously discussed in relation to social presence, on-line behavioural measures have the potential to supplement subjective self-report, and these further investigations, with the additional online behavioural measures, could enrich the understanding of what is driving perceptions of rapport during interactions with agents and avatars in virtual worlds. Manipulating the user's beliefs about whether they are interacting with a person or a computer would also be interesting in that it could uncover differences between how a user will respond to another human compared with a computer, in terms of behaviours towards and perceptions of the interlocutor. Another point of note is their report that a human speaker looks at a listener approximately 65% of an interaction. As previously reported, Kendon's studies found a figure of between 28% and 70% (Kendon, 1967), and as previously stated, Argyle and Dean (2013) found that a listener will look at a speaker during

approximately 75% of a dyadic interaction. These are important points in relation to the experimental design section in this thesis (in particular in the human-agent experiment – see the ‘Gaze’ section of *Independent Variables* in the Method section of experiment 2). In both Kendon’s and Wang and Gratch’s research, the interaction involved having a conversation or simply talking to a conversational partner. Could it be that the amount of visual attention that one gives to a partner is moderated by the type of interaction? Would a task-based interaction, for example, result in different amounts of looking, compared with a conversation recounting a story?

In a further study, the rapport agent’s gaze and gender were manipulated in order to investigate evaluations of gendered humans’ behaviours (Kulms, Krämer, Gratch, & Kang, 2011). They found that female agents were evaluated more positively and that gaze avoidance was evaluated negatively. It would seem, therefore, that somewhere between staring and gaze avoidance would be appropriate for effective communication. The question is, therefore, where between these two extremes?

In an investigation into the relationship between mutual gaze and social presence, Shahid and colleagues looked at how children’s perception of social presence varied during game playing via 3 different communicative settings: face-to-face, video-mediated with no access to mutual gaze and video-mediated with access to mutual gaze (Shahid, Krahmer, & Swerts, 2012). They found that access to mutual gaze increased perceptions of social presence. The study does not, however, measure mutual gaze per se, but only establishes that if children have the potential to engage in mutual gaze that they will experience higher levels of social presence. Clearly the next step in this situation would be to find out how much mutual gaze actually occurs during the interaction, and whether increased levels of mutual gaze are related to increased perceptions of social presence, or whether it’s just the knowledge of access to mutual gaze. The authors report that gaze behaviour varies as a function of age, so it would be interesting to see if the relationship between access to mutual gaze and social presence perceptions hold for adults. If such a relationship was found, along with a relationship between social presence perceptions and the *uptake* of mutual gaze, then it opens up the potential for using both high- and low-level behavioural

measures (mutual gaze and task performance) to supplement existing methods of self-report social presence measurement.

Bente et al. (2007) aimed to demonstrate that increasing the duration of gaze towards an interlocutor (i.e. maximising opportunities for mutual gaze) results in more positive social perceptions of a conversational partner. Dyads communicated in a computer-mediated interaction via avatars whose eye movements were programmed to vary across conditions. Avatars “offered eye contact” for either 2 seconds or 4 seconds, followed by 2 seconds of gaze aversion (p. 209). The long gaze condition was found to produce more favourable judgements, including higher social presence ratings. In a second experiment, a further two looking conditions were included, resulting in the avatar offering eye contact for either 2s, 4s, 8s or 16s. They aimed to investigate the effects of gaze duration on perception of personality traits and social presence ratings. They also included the variable *gender*, by using mixed-sex dyads. They found that the longest gaze duration (16s) resulted in the most unfavourable perceptions, and the lower gaze conditions (2s and 4s) elicited judgements of the interlocutor being more active and assertive. Finally, females were more sensitive to their male partner’s variations in gaze than vice versa.

This is a useful methodology for investigating the effects of gaze on person perception, as well as on an interlocutor’s gaze. In neither of these experiments were the interactions task-focused, as such. In the first study, participants were asked to chat to get to know each other, and in the second, they were told to “chat to form a detailed impression of the partner” (p. 210). This type of paradigm would be useful for investigating effects of looking behaviour on eye movements, person perception (social presence) *and* task performance.

## **2.5 From avatars to embodied conversational agents**

With the advances in computer technology, there has been an increase in the amount of collaborative interaction between humans and computers, often with the assistance of embodied conversational agents (ECAs). ECAs are “computer-generated cartoon-like characters that demonstrate many of the same properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and

nonverbal communication.” (Cassell, Sullivan, & Prevost, 1999). To date, evaluations of interactive systems that use ECAs have focused mainly on self-report measures from users, with a few notable exceptions. Cassell & Thórisson (1999) looked at participants’ evaluations of combinations of content, emotional and envelope feedback in animated conversational agents, and compared these with videos of behaviour and speech during interaction with the agents. Emotional feedback refers to facial expressions indicating emotional responses, such as surprise or happiness, and envelope refers to feedback which is directly related to the process of the conversation, such as nodding in agreement. Content feedback consists of speech or actions in response to questions or requests for information, but involves no emotional or envelope feedback. Envelope feedback was found to be the most important during interaction, more so than emotional feedback, in terms of subjective perceptions and efficiency of the interaction. Another example is Bailenson and colleagues’ previously discussed comparison between behavioural (interpersonal distance) and self-report measures of copresence in immersive virtual environments, where they found a disparity between the self-report measures of likeability, copresence, status and interest and the physical space given to different types of agent (Bailenson et al., 2004).

One area where ECAs are widely used is in virtual learning environments. In a review by Johnson and colleagues several agents were discussed, including Steve (Soar Training Expert for Virtual Environments; Johnson, Rickel, & Lester, 2000; Rickel & Johnson, 1998). Steve assists the user in learning how to carry out procedural tasks in an immersive virtual environment that represents part of a US Navy surface ship. This agent demonstrates to the learner how to perform procedural tasks in an IVE, such as checking a dipstick, which the student can then imitate, meaning that they will be more prepared when it comes to working on-board a real ship. This reduces the potential costs involved in training, financial as well as expert time, as well as reducing the potential for accidents before the user is properly trained. Steve is particularly helpful in teaching learners how to physically perform procedural tasks in a rich immersive virtual environment. One disadvantage of the Steve system is that it is not easily generalisable to domains other than teaching procedural tasks on-board a navy ship, due to the complex nature of the IVE.

Adele, in contrast, is an application designed for running over the Web (Johnson, Shaw, Marshall, & LaBore, 2003; Shaw, Johnson, & Ganeshan, 1999). Here, the student is assisted in learning knowledge, rather than how to carry out procedures, such as teaching medical and dental students about, as well as testing them on, the order of procedures that should be followed when presented with particular symptoms. One of the main benefits of this system is clearly the opportunity for the student to learn and practise without possible danger to patients' lives. Because this system does not require an immersive 3D environment, Adele is more generalisable to domains other than medical and dental training. Some kind of 3D virtual environment that has the versatility of Adele, while retaining the immersive virtual qualities of Steve would be an ideal one in which to study interaction with ECAs.

One aspect of the Steve system that appears to have been overlooked is the formal evaluation of the system. Although the authors say in several of their papers that evaluation of Steve is forthcoming and that continuing informal evaluation is being carried out, there appears to be none reported (Johnson & Rickel, 1997; Rickel & Johnson, 1998). This is one area that Adele is clearly superior, as evaluation has been carried out on over 250 students over 5 years (Johnson et al., 2003; Shaw et al., 1999). This not only confirmed students' liking for and perceived usefulness of the system, but also uncovered aspects that may not have been discovered without formal evaluation, such as Adele's perceived realism. In terms of evaluation of systems involving ECAs, much has concentrated on subjective evaluations, with the exceptions mentioned above.

More recent research has looked at how to design agents for effective human-computer interaction. Andrist and colleagues analysed video data of dyadic human-human conversations, paying specific attention to gaze aversion (Andrist, Mutlu, & Gleicher, 2010). The aim was to use the gaze behaviour from the dyads as a model for the behaviour of virtual agents. They tested the effect of the behaviour of the resulting agents on human participants and found that when employing gaze aversion, they were deemed to be thinking, were able to elicit more disclosure from the participants as well as regulating turn-taking in conversation.

Avatars are similar in appearance and behaviour to agents, but where ECAs are controlled by computers, avatars are driven by humans. Avatars are widely used during computer mediated communication, especially in virtual environments such as Second Life (<http://secondlife.com/>). As previously described, the environment is such that it can be adapted to different domains, so that the user can carry out different tasks whilst remaining in the same 3D virtual environment. This means that it is possible to import a given task, such as teaching a method or a procedure, and evaluate the method by recording the screen and analysing events post-procedure. Indeed, studies have been carried out within the environment, covering varying areas of research. Yuen (2013), for example, looked at the use of Second Life as a tool for delivering treatment for Social Anxiety Disorder. They found significant improvements in symptoms, such as depression and anxiety, and that the effect sizes for these improvements were comparable to those relating to face-to-face treatments. Yee and colleagues' (2007) study into the persistence of nonverbal social norms in Second Life has already been discussed and Kraemer (2013) found evidence for group influence within Second Life.

As can be seen from the few studies discussed here, Second Life is an ideal platform for the purposes of studying many aspects of social interaction, as it provides the 3D environment of Steve, whilst retaining the adaptability of Adele. Another, more practical reason for using SL is that it is freely available for public use. The final and most important reason for using SL, in relation to previous systems using ECAs, is that it is possible to program an avatar to act as a robot within SL. It is further possible to manipulate the robot to say or do almost anything that the experimenter desires. This means that it is a useful platform for developing an experimental paradigm for investigating interaction with ECAs as it will be adaptable to different domains.

## **2.6 Eye Tracking and ECAs**

Recent years have seen an increase in the amount of research establishing the validity of the use of eye tracking as a usability technique in human computer interaction. Some such studies have demonstrated the importance of eye tracking as a usability tool, as well as producing methodologies that can be widely used in the evaluation of



different types of interface (McCarthy, Sasse, & Riegelsberger, 2003; Pan et al., 2004). In terms of using eye tracking to investigate the use of ECAs, Prendinger et al. compared the eye movements of subjects viewing the presentation of an apartment by 3 different types of media – an animated agent, a text box and speech only (Prendinger, Ma, Yingzi, Nakasone, & Ishizuka, 2005). The participants were asked to fill in a questionnaire relating to their general impression of the presentation and their ability to recall specific items. The questions of interest in this study were (1) Is the user paying attention to the interface agent (2) To which part of the agent is the user attending and (3) Can the agent's verbal or gestural behaviour direct the user's focus of attention? Here the agent can gesture left and right. The authors found, among other things, that an agent's referential (arm or facial) gestures may direct the attention of the user to the intended referent better than the voice and that users often redirect their attention back and forth between the animated agent and the reference object. The interface in this study does not involve active interaction. The user passively views the scene, and the interaction is not task-driven. Although the authors argue that "users watching a presentation interact – even involuntarily – by their eye movement activity", it is possible that using a more interactive task may be more informative about how the viewer uses information presented by the agent (Prendinger et al., 2005; p.3). It would be interesting to find out if the same results would be found if the interaction were more task-driven, and therefore a comparison between eye movements, subjective responses and task performance measures could be made.

Witkowski et al. (2001) investigated how useful it is to study eye movements for the evaluation of character agents. Customers view an interface, MAPPA (Multimedia Access through Personal Persistent Agents) that represents an online wine store. The images are presented to them using a method similar to the *Rapid Serial Visual Presentation* (RSVP) method, which means that a large array of products can be presented on a small screen. The authors here state that although many systems involving ECAs are in development, or in general use, empirical evaluations of how the user reacts to the characters are not very common. The authors wished to examine the relationship between the appearance of the agent and the user's eye movements. They wanted to establish if there were individual differences between

participants, if any patterns emerged over repeated exposures to the agent and if there was any relationship between eye movement data and subjective comments about the agent. Here is one of the rare instances of the relationship between subjective self-report and eye tracking measures being explored. The procedure in this study involves registering with the MAPPA system and viewing introductory help. This is followed by an opportunity for the user to explore the interface, prompted by a facilitator to try out various aspects of the system, such as using the *left* and *right* buttons, bringing up at least one product dialogue box and using the purchase dialogue box to add an item to the basket. The study did not, again, involve a uniform task being carried out by all of the participants, with each individual carrying out different procedures from which the data evolved. The dependent variables included number of fixations, as well as the overall time, mean time and standard deviations of the fixations. The results (from the 5-participant pilot study) suggested that for the majority of time participants were looking at the text box above the agent, more so than the time spent looking at the agent. This suggests that the text itself commands more attention from viewers of the interface than the agent. This finding could be taken into account when designing interfaces – is it better, for example, to present information verbally rather than in text format? Less time was spent looking at the product and other interface objects than either the text or agent. The authors make clear that the study was only intended as a pilot study, one that could establish the use of eye tracking as a valid method for evaluation of the effects of agents in a user interface.

In an unpublished study by a University of Edinburgh Human-Computer Interaction MSc group, a comparison between self-report and eye tracking measures was made for individuals while they were being guided around a map from a set start to finish point by ‘Genie’, an ECA acting as a tour guide (Dalziel-Job, Papananagiotou, Fourneron, Vattoly, & Tajwer, 2006). These results were compared with self-report and eye-tracking measures resulting from the trials where instructions were given by voice alone. It was found that the average total amount of time spent gazing at the agent was only 0.12%. Despite this low figure, however, 75% of the users found Genie to be a more helpful tour guide than voice alone. This is an interesting finding, as it could be expected that individuals finding an ECA helpful would be likely to

utilise it, i.e. look at it more. One possible reason for these results could be that the task itself was difficult, requiring high cognitive load, and the agent, although pointing in the correct direction, was not critical to the task; all required information could be taken from the voice alone. When Genie was available, however, the proportion of correct locations visited was more than when voice alone was available, suggesting that, despite not looking at Genie, its presence helps in task performance. It must be noted, however, that no significance tests were carried out on these results, so any patterns that were uncovered can only be treated as indicative of a possible trend.

A possible extension of this study could be to ensure that the agent holds task-critical visual information in one condition so that the user is forced to look in order to successfully complete the task (as with the Steve system where information critical to carrying out the procedural task was demonstrated by the agent). Would this increase or decrease the perceived helpfulness of, or amount of gaze towards, the agent?

Much of the previous research into eye movements in SL has been dedicated to using eyes to control a user's avatar, a method especially valuable for individuals with disabilities that inhibit them from using a standard mouse and keyboard (Vickers, Istance, Hyrskykari, Ali, & Bates, 2008). Dalziel-Job, Nicol and Oberlander (2008), however, recorded users' eye movements during a task-orientated interaction with a programmed avatar (agent) to investigate how individuals respond to informative compared with redundant gestures in SL.

## ***2.7 Avatars, agents and social interaction***

As previously discussed, the advances in computer-mediated communication and human-computer interaction have resulted in an increase in research into how people interact with or via computers. The use of computers and virtual environments to study social interaction has clear benefits, from high levels of experimental control to the ability to replicate experiments near-perfectly (Blascovich et al., 2002). One of the questions about using agents or avatars to study social interaction is how much of the behaviour during human-computer or computer-mediated communication can be

generalised to face-to-face interaction? Or conversely, how much of the current theory about face-to-face communication can be generalised to mediated or human-computer behaviour?

Since much research into social interaction is lab-based, due to the ability to replicate the paradigms and the high levels of experimental control, it is important to establish how much the results of these studies carry over to real-world interactions. One of the common methods by which researchers have studied how people view faces is by presenting participants with videos and recording their eye movements. Blagrove et al. (2012) identified the lack of recent research into how people allocate attention to faces in real-life situations; the majority of findings in the area of how people view faces in a lab situation. They compared the eye movements of people during task-based real-world interactions (such as buying a coffee) and participants watching videos of the real-world interactions, from a first-person view (i.e. from the view of the participants in the real-world condition). More fixations were made to the task-relevant faces (e.g. baristas) in the lab-based viewings than in the real-world interactions. They also found that the lab-based participants spent a higher proportion of the time looking at the task-relevant faces than bystanders' faces, compared with real-world participants. These differences could be due to the tasks involved; participants in the video condition had to merely watch what was happening, whereas real-world participants had to complete tasks (for example, buy a coffee, navigate an environment), which could compete for their visual attention.

Foulsham et al (2011) also compared allocation of visual attention in real-world situations and during the viewing of the videos of the same situations. One of their findings was that real-world participants were less likely to fixate on the faces of other people who were close to them than viewers of the videos were. These two studies are interesting in showing that viewers of videos can behave in a different manner towards social stimuli than people in real-life interactions, meaning that one cannot necessarily generalise results from lab-based video studies into social attention to the real world. One of the obvious issues with viewing videos, rather than real-life, is that although there are social cues available, it cannot be said to be an interaction; the viewing is passive and one-way. Freeth, Foulsham and Kingstone

(2013) compared eye movements of interviewees under two levels of social presence. Participants were asked four questions either in a face-to-face interview or by pre-recorded video (this was between-participants) and the interviewer either looked directly at the interviewee or away during the questions (within-participants). During the face-to-face interaction it was found that participants spent more time looking at the experimenter's face than body when being looked at directly, and that these effects were stronger during answers than when questions were being asked. In the videotaped sessions, however, there was no effect of interviewer-looking on where the participant looked, indicating that studying how people view faces and eyes by how they view videos is not comparable to face-to-face interactions. These results demonstrate that, in studying social behaviour, it is important that some sort of interaction is involved. The ideal, therefore, would seem to be interaction through avatars or agents, where social interaction is involved, but experimental control and replicability remains high. The question, then, is how do findings from such studies relate to face-to-face interactions?

Clearly the levels of social presence will be lower in mediated communication than face-to-face, and presumably there will be likely to be fewer non-verbal cues, but are people's social responses to computer-generated agents and avatars similar to those in face-to-face interactions and vice-versa? We have already seen from Yee and colleagues' research that there is evidence to support the persistence of non-verbal social norms during computer mediated communication within virtual environments (Yee et al., 2007). There are several studies which show that people respond to and treat computers in a social manner, similar to the way in which they treat humans. For example, Nass and colleagues (1999) found that people responded more politely to a computer when it asked for feedback about its own performance than when asked by another computer, as well as finding that computers were able demonstrate personality with only minimal manipulation (Nass, Moon, Fogg, Reeves, & Dryer, 1995).

Kim and Sundar (2012) investigated why people treat computers like humans. They argued that it could be due to a mindless tendency, rather than a mindful (i.e. conscious) anthropomorphism, because of the human-like bonds that we have

developed over long-term use with them. By manipulating the presence/absence of a human-like agent as well as the level of interactivity (low/high) they found evidence for mindless anthropomorphism during interaction with computers. We can see, therefore, that humans attribute human-like qualities to computers and as a result will respond to them, to some extent, in a social manner as they do other humans.

Krämer et al (2012) looked at whether human-human theories can be adapted for human-robot / human-agent interaction, or if new theories need to be developed. They concluded that “people readily engage in communicative behaviour as they know it from face-to-face behaviour as soon as something appears to be sufficiently social” (p.232). On the one hand, the agency assumption says that the social influence of human-controlled avatars will always be higher than that of agents, and that believing that one is interacting with another human is enough to elicit social behaviour towards an interlocutor. Conversely, it could be that human-like behaviour will automatically provoke social reactions, even if the user knows that their interlocutor is a computer (Nass, Steuer, Tauber, & Reeder, 1993). Appel et al. (2012) investigated this by manipulating the level of agency (agent/avatar) and level of social cues (low: textchat / high: virtual human). They found that participants in the virtual human condition, compared with the text-chat condition, attributed more positive characteristics and a stronger sense of mutual awareness to their conversational partner, as well as allocating more attention, supporting the idea of behavioural influence: more human-like behaviour (i.e. more social cues) elicits social reactions. They did, however, report higher levels of social presence for avatars than for agents, which adds weight to the agency assumption: that the social influence of avatars is always going to be higher than that of agents. Blascovich and colleagues propose a threshold model of social influence, incorporating both the agency assumption and the concept of behavioural influence (Blascovich et al., 2002). According to this model, a computer agent or avatar must have to have the right combination of social presence and/or behavioural realism to reach or surpass the threshold for social influence upon an interlocutor. They define behavioural realism as “the degree to which virtual humans and other objects within IVEs behave as they would in the physical world” (p.111). If social presence is high (e.g. if a user is interacting with a human-controlled avatar), then behavioural realism can be low

(the example they use is an avatar consisting of a smiley face on a beach ball) and the threshold is still met. For a computer-controlled agent to exert social influence then behavioural realism must be high in order to compensate for the low level of social presence. As levels of behavioural realism and social presence both increase above the threshold, stronger social influence effects will be seen. According to Blascovich and colleagues, this model holds for high level responses, such as meaningful conversation and joint tasks, but not for low level responses or less consciously controlled processes (e.g. eye movements).

In terms of Appel et al (2012), the finding that there was no difference between the amount of attention allocated to the agents and avatars, but the avatars were perceived as having more social presence would support the threshold model of social influence; high level responses (social presence reports) are being affected by manipulation of agency, but not low-level responses (attention allocated to interlocutor). In relation to this thesis, making a participant believe that they were interacting with either a person or a computer should, in theory, vary the perception of social presence of the interlocutor. For the purposes of the current study, this is a way in which the beliefs of the user could be manipulated without altering the richness of social cues available to them.

## ***2.8 Social facilitation and inhibition***

Social facilitation is one of the oldest areas of social psychology research. The concept of social facilitation was reportedly first identified by Triplett (1898), after he noticed that the presence of others affected an individual's performance during cycling. Zajonc (1965) reviewed research into how the presence of others affected an individual's performance and concluded that many of the results were contradictory, with some finding performance improved and others finding that the presence of others interfered with performance. For example, Bergum & Lehr (1963) looked at the performance by National Guard members during the monitoring of light panels. The task was to signal whenever a light failed to go on in the correct sequence. They were trained for 20 minutes and then monitored the panels for 135 minutes. One group performed alone and the other was told that a more senior member of staff would visit periodically to observe their performance. The supervised participants

performed better than the alone participants (20% of light failures were missed in the supervised condition, compared with 64% when alone).

Dashiell (1930) also found improved performance in the presence of others, this time on simple multiplication or word association tasks. Pessin (1933), however, found that participants were slower and less accurate at learning nonsense syllables in the presence of another person than when on their own. Husband (1931) found that the presence of another person interfered with learning of a finger maze, in terms of the time taken and accuracy. Zajonc (1965) suggests that this discrepancy is due to the nature of the dominant response in the specific tasks. When one is learning a task, the dominant response (i.e. the one that occurs most often) is the wrong one. When a task has been learnt, the dominant response is the correct one. Zajonc's conclusion was that the presence of an audience facilitates the dominant response on any given task, due to arousal caused by the presence of others. Therefore, if the task in question is not a familiar or learnt one, then the presence of an audience will result in more incorrect responses. If the task is a learnt one, then the presence of an audience will result in improved performance, or the increase in the number of dominant responses: the correct responses.

The question, in relation to this thesis is: do these facilitation/inhibition effects carry over to virtual environments, and if so, is there a difference between the effects seen during interactions with agents and avatars?

Zanbaka and colleagues (2007) examined whether social facilitation/inhibition effects were the same when completing easy or difficult maths tasks in the presence of a human observer or a computer-controlled agent. In the agent conditions, the virtual observer was either projected onto a wall in the room or the user was immersed in an immersive virtual environment with a virtual human seated next to them within the environment. They found that performance on complex tasks was inhibited in both the human observer condition and the virtual agent conditions. They found no facilitation effects in either the real or virtual human conditions. They suggest that the lack of facilitation effect may have been due to a ceiling effect on the task performance. They conclude that "many of the rules that apply in human-human interaction carry over to interacting with interface agents and computers" and that



social interaction theories should be taken into account when developing interface agents (Zanbaka et al., 2007; p.8). The next question would be: can social facilitation/inhibition effects be found during computer-mediated communication, i.e. during interaction with a human-controlled avatar, and are the effects similar to those during interaction with a computer-controlled agent? If so, this could be a useful way of assessing the social influence that virtual characters have.

Blascovich et al. (2002) report a study in which they investigated social facilitation within immersive virtual environments (this has been previously alluded to in relation to the threshold model of social influence). Participants performed either well-learned or novel tasks within an IVE, either alone or in the presence of a male and female observer. They were told that the observers were either avatars or agents (although they were always computer-controlled agents). No difference in performance was found between the agent and alone conditions, but there was evidence of social inhibition in the avatar condition, in that participants performed worse on the novel tasks compared with the alone condition. They did not find any social facilitation effects, however, with no differences in performance emerging between the alone and avatar conditions for the well-learned task. They report that this could (as suggested by Zanbaka and colleagues) be due to a ceiling effect due to the learning phase of the tasks. If this study could be replicated without the ceiling effect, it could help discover whether social facilitation does indeed occur within virtual environments. What is clear is the usefulness of task performance as an indicator of an individual's social perception of an interlocutor, meaning that potentially it could be included, along with gaze measures, in a model of social presence, identifying facets of SP that may not be detected by questionnaires alone. In fact there is already some evidence of a relationship between co-presence ratings and task performance in a previously mentioned study, although the correlation did not reach significance (Bailenson et al., 2005).

Abeele and colleagues report that social facilitation theory “offers interesting opportunities for further clarifying the concept of social presence, in particular the role of perceptual awareness as a necessary prerequisite of social presence” (Abeele, Roe, & Pandelaere, 2007). Another point to be made here is that if the threshold

model of social influence is correct, then in any interaction with an agent or avatar, social facilitation/inhibition will only occur if the threshold for social influence has been met. An interlocutor must either be (or be believed to be) an avatar, or the behaviour must be sufficiently behaviourally realistic to wield social influence. Therefore the concepts of social facilitation/inhibition and the threshold model of social influence are inextricably linked when discussing interactions within virtual environments.

## **2.9 Research questions**

It can be seen from the aforementioned literature that there is some evidence of relationships between the three measures: social presence, mutual gaze and task performance, although they have not, to date, been drawn together to find out how they all relate to each other within a task-oriented computer-mediated interaction. In order to do this, an appropriate experimental paradigm was needed. The next chapter introduces a human-human study into the relationship between mutual gaze and task performance within Second Life. The two main research questions relating to the human-human study concern mutual gaze and task performance and will be addressed in chapter 3:

1. Does constant staring by one individual at another maximise mutual gaze between the pair?
  - Is constant staring by one conversational partner at another an appropriate method for mutual gaze research, or could it introduce an experimental artefact, resulting in the suppression of mutual gaze?
2. Does mutual gaze facilitate task performance within this paradigm?

The human-agent study, covered in chapter 4, will address these additional research question:

3. What is the optimum amount of looking by one individual at another for maximising:
  - Mutual gaze
  - Task performance

- Perceived social presence?
4. What is the relationship between purported agency (manipulated social presence), mutual gaze, task performance and perceived social presence?
    - Can behavioural measures in this paradigm (gaze/task performance) identify variations in actual social presence which may go undetected by a self-report questionnaire?
  5. What is the function of gaze, specifically mutual gaze, during task-based interactions within Second Life?

It is anticipated that the answers to the above questions may present further questions, which will be addressed as they arise. The first two research questions will be addressed using a novel human-human paradigm that enables the recording of two users' eye movements during a task-based interaction within Second Life. Chapter 3 describes the methods used within this paradigm.

### **3 Chapter 3: Experiment 1 – Human-Human Study**

The human-human study investigates the effect of one person staring at another on the amount of mutual gaze between the dyad, during a task-orientated interaction. It also looks at the relationship between mutual gaze and task performance. It investigates whether the method of having one individual stare continuously at the other, as utilised by Bailenson et al., is the most appropriate one for studying mutual gaze, or if this looking behaviour, in fact, decreases the amount of mutual gaze between the dyad, thus introducing an experimental artefact (Bailenson et al., 2002)? Furthermore, it is of interest to establish if mutual gaze can facilitate performance on computer-mediated arithmetic tasks, as Fullwood & Doherty-Sneddon found during recall tasks using VMC (Fullwood & Doherty-Sneddon, 2006). If this is the case, then the more mutual gaze there is between a dyad during a task-orientated interaction, the better the score will be on the task.

#### **3.1 Method**

This section introduces the methods used to examine the relationship between staring, mutual gaze and task performance in the human-human study. These methods will be elaborated and extended in chapter 4 for use in the second experiment (see the human-agent Method section). Following description of the methods used for both studies, we will re-visit the experimental designs in chapter 5, providing supplementary material concerning the motivations behind design decisions, as well as challenges encountered during the process of data collection (see Chapter 5: Experimental Design). The aim of chapter 5 is to focus on the methodological aim and how it was achieved.

##### **3.1.1 Independent Variables**

To find out whether constant staring by one conversational partner at another results in increased mutual gaze between the dyad, participants completed tasks under two looking conditions – *staring* and *not-staring*. In the staring condition, the instruction giver's (IG's) avatar stared constantly at the instruction follower (IF) during the interaction, and in the not-staring condition, the IG's avatar looked at the IF intermittently.

### **3.1.2 Dependent Variables**

The first dependent measure is the amount of time the instruction follower spends looking at the eyes of the IG's avatar under the 2 different conditions – staring and not-staring. During the staring condition, this also amounts to mutual gaze between the dyad, since at all times the IG's avatar is staring at the follower. In the not-staring condition, however, the time during which the IF looks at the IG's avatar's eyes can constitute either mutual gaze – the IG's avatar is looking at the IF's eyes – or monitoring gaze – the IG's avatar is looking at something else. After establishing how much of the trial under each condition the IF spent looking at the IG's avatar's eyes, this time was split into mutual gaze time and monitoring gaze time. In order to establish the effect of staring on the amount of mutual gaze between a dyad, the absolute amount of mutual gaze in the 2 conditions was measured, as well the proportion of opportunities for mutual gaze that were taken up. This was defined, in the not-staring condition, as the amount of time during which the IG's avatar was looking at the follower that the IF was looking back. In the staring condition this was the same as the absolute amount of mutual gaze.

The second dependent measure was task performance. The initial measure of this was how many of the 15 tasks in each block the IF answered correctly. If task performance is indeed facilitated by mutual gaze, then more mutual gaze between a dyad will lead to increased task performance.

### **3.1.3 Participants**

52 participants (mean age 23.4; 25 male) were recruited via SAGE (Edinburgh University's employment website) and by word of mouth. Each participant was a volunteer and received £5 financial remuneration for participating in the study. Sessions lasted between 30 and 45 minutes. Each participant passed an Ishihara Test for colour-blindness (Ishihara & Force, 1943).

### **3.1.4 Apparatus**

Participants viewed the experiment on a 19 inch CRT display, sitting approximately 42cm away from it. The instruction-follower was free to interact with a standard

mouse, but asked not to touch the keyboard. The instruction-giver was asked not to touch the keyboard or mouse in the first condition, but to use the mouse to hover the cursor over the tile that he or she was describing in the second condition. Each participant was linked up to an SR-Research EyeLink II head-mounted tracking system. Additionally, the instruction-follower wore a set of headphones and the instruction-giver wore a microphone headset to enable the follower to hear the giver's instructions via Second Life. The volume at which the instructions were heard could be adjusted to suit the instruction-follower. The procedure took place within a sound-proofed room. The sample rate was set at 500Hz and the participants' dominant eye was tracked monocularly. A 9-point calibration matrix was used at the start of each participant's experiment and between blocks if needed. Camtasia Studio (produced by TechSmith) recorded what each participant could see on the screen throughout the procedure. It also incorporated what the instruction-follower could hear, as well as what the instruction-giver said, throughout the experiment into the movie files that were generated to allow for easier analysis.

### **3.1.5 Design**

In a within-subjects design, all participants carried out 15 tasks under 2 conditions – the staring and the not-staring conditions. The order of the conditions remained constant for each dyad, but the tasks themselves were counterbalanced between the participants for the 2 conditions to control for effects due to task itself. The 2 conditions were as follows:

The instruction-giver's avatar looks directly at the instruction-follower, providing a staring condition. This was achieved by asking the IG not to move the mouse at all during the first block, resulting in the default behaviour of an avatar in Second Life – staring straight ahead – i.e. at the follower.

The IG's avatar looks at the tiles as the IG describes them, thus providing a non-staring condition. This condition was achieved by asking the IG to move the cursor so that it hovered over the tile that they were describing. This automatically moves the IG's avatar's gaze to the said tile.

It is important to note at this stage that, although the IG-avatar looking is informative in the not-staring condition, it is redundant, since the follower does not need to look

where the giver's avatar is looking in order to complete the task; all information needed is provided verbally. In both conditions, the IG was allowed to give instructions as he or she wished. Since the comparison to be made was between the 2 conditions – i.e. a related design (within subjects) – it was of no interest, at this stage, as to the content of the instructions, only a comparison between the behaviours in block one and block two.

### **3.1.6 Stimuli**

A building comprising of 1 large closed room was built on VUE, the Edinburgh University Island within Second Life (see <http://vue.ed.ac.uk/>). Within the room there were 2 chairs facing each other with a glass screen between them. These chairs were for the participants' avatars to sit on. In front of each chair was a panel which was only visible to the person whose avatar sat in that chair and was hidden from the other participant. On the IG's side there was a panel containing the instructions to be imparted to the IF (see Figure 1). These took the form of 5 Instruction Tiles making up a short arithmetic task: 3 numbers separated by 2 arithmetic operators (plus or minus) in black text on a white background. On the IF's side were 3 tiles on which were presented 3 possible answers, the Response Tiles, again black numbers on a white background (see Figure 2). On the glass screen between the 2 avatars were 7 Stimulus Tiles, which were visible to both participants. Each stimulus tile had a white number on a background of a particular shape and colour. These were either red, green or blue and a circle, square or diamond shape. There was no more than one tile with each specific colour-shape combination and no more than one of each number on view at any one time.

Each chair was programmed so that when an avatar was sat upon it, the camera view – i.e. what the participant controlling the avatar could see – was taken over. The camera view for each avatar was such that each participant saw through their avatar's eyes, and the view was such that it maximised the view of the other avatar, the stimulus tiles and their own tiles in front of them – instruction tiles or response tiles. The purpose of sitting each participant's avatar on a chair was that each participant had exactly the same view of the experiment, and this also had the benefit that the

avatars were unable to walk around the room, which would have disrupted the experiment, although they were free to do so after the experiment if requested.

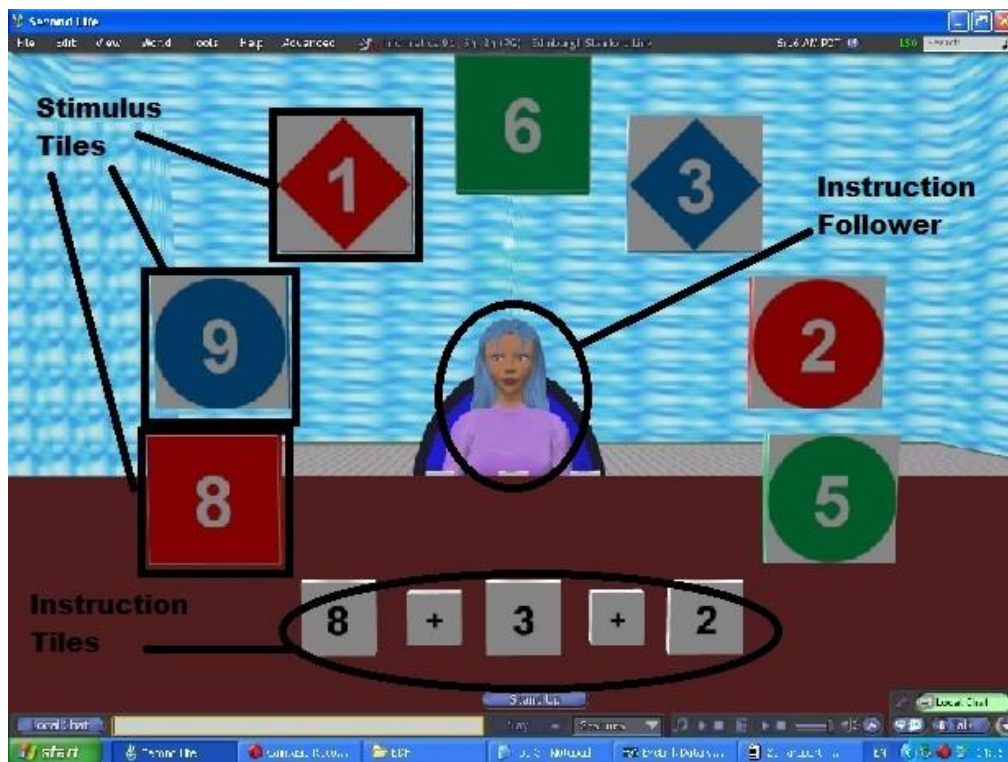


Figure 1: Instruction Giver's view of Instruction Follower's avatar.





**Figure 2: Instruction Follower's view of Instruction Giver's avatar**

Each block consisted of 15 tasks. Each task was a short, relatively simple, arithmetic problem. Each problem was presented to the instruction giver via the instruction tiles, who then had to relate it to the instruction follower without naming the numbers involved. In the IG's view above for example, the IG would say something along the lines of "red square plus blue diamond plus green circle" and the IF, in response, would click on the left-most response tile, indicating that 13 was the correct answer (see **Figure 1** and **Figure 2**). Upon the IF clicking on one of the response tiles, all of the texture on the tiles would change for the next task. Also in the room was a camouflaged screen behind which Tracker, a bot controlling the stimulus texture changes, was hidden from view. Neither participant could see, or was aware, that Tracker was also in the room.

Textures (numbers with or without shape/colour backgrounds) were imported into Second Life to create all of the tiles – the instruction, stimulus and response tiles - and attached as a texture to the tile as required, controlled by the Tracker program. The textures for the tiles (numbers / shapes / colours) were created using Microsoft

Paint version 5.1 and the GIMP (GNU Image Manipulation Program) before being imported into Second Life.

### **3.1.7 Procedure**

Participants arrived at the Joint Eye-tracking Laboratory at Edinburgh University in pairs and were randomly assigned to Instruction-Giver and Instruction-Follower roles. Each was seated in front of a computer and asked to read and sign a consent form (see Appendix A). They were then given a short set of verbal instructions, which gave them both a brief overview of what the experiment consisted of. These instructions included telling the IG not to touch the keyboard, and how to convey the sum to the IF (i.e. without using the names of the numbers themselves). The IG was asked to speak as clearly as possible and to only give the instructions once. The experimenter checked that both participants understood the tasks and what they had to do. To produce the looking condition manipulation, before initiation of the first block of tasks the IG was instructed not to touch the keyboard, only to give verbal instructions. This resulted in the staring condition. Before the 2<sup>nd</sup> block of tasks began, the IG was instructed to use the mouse to hover the cursor over the tile that they were describing, producing the not-staring condition.

Each participant was set up separately. For each participant Camtasia screen recorder was initialised. It was then determined which was the subject's dominant eye, and the eye tracker was placed onto the participant's head. It was explained that, while the head set with the eye tracker had to be secure enough not to move, it must still be comfortable, and that it would be tightened until the participant felt that it was tight enough. Following this, Second Life was started up on both PCs, and both avatars were logged in. Once SL, Camtasia and the hardware had been set up, both participants were led through the 9-point calibration process. After this, the program controlling Tracker Merlin, and the texture-changes on the tiles, was initialised for one of 2 blocks of tasks – tasks A or Tasks B. Each of these 2 contained 15 tasks – 30 in all – and the two sets of tasks were counterbalanced between all of the experimental runs to control for possible artefacts due to materials. Upon completion of both trials, a written post-test questionnaire was administered for the participant to fill in independently (see Appendix B). The function of the

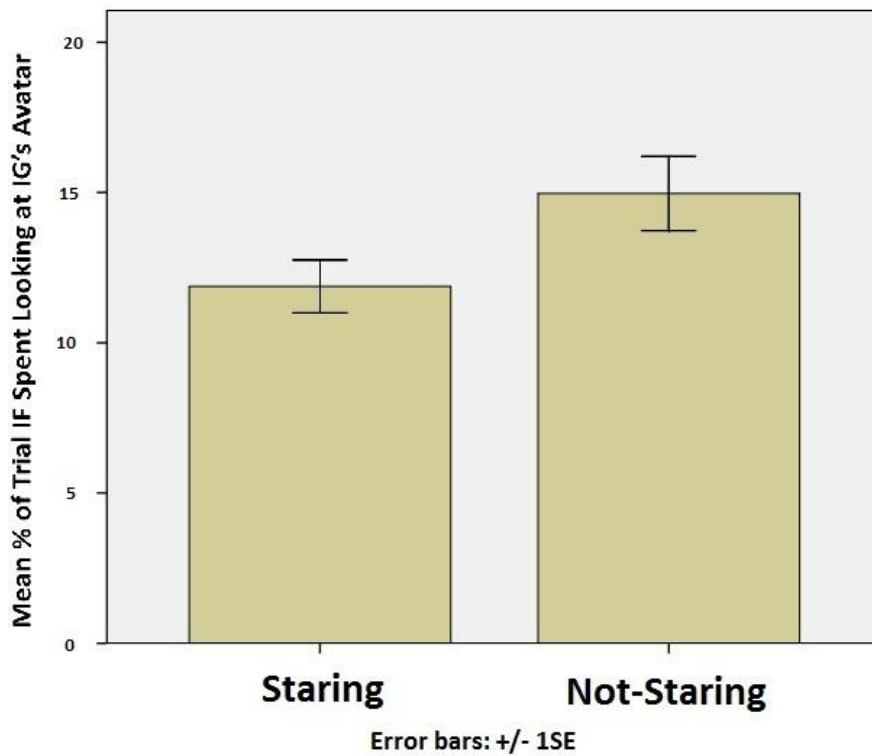
questionnaire was to collect demographic information. The participants were then asked to sign a receipt and were given £5. After this, the purpose of the study was disclosed, as well as a brief tour of SL if requested.

## **3.2 Results**

### **3.2.1 Hypothesis 1 – staring and mutual gaze**

It was anticipated that in the *staring* condition, the instruction follower would prefer not to engage in mutual gaze, as such unnatural looking behaviour by the instruction giver will dissuade the follower from engaging in eye contact. It was hypothesised, therefore, that staring would negatively impact on the amount of mutual gaze between a dyad, and if that was the case, then the method utilised by (among others) Bailenson et al. should take into account the potential impact of such an experimental manipulation.

To establish a baseline for each condition, the amount of attention the IG attracted in each looking condition was compared; was there a difference between the amount of overall time spent looking at the IG's avatar in the agent and avatar conditions, regardless of whether the IG was looking at the IF?

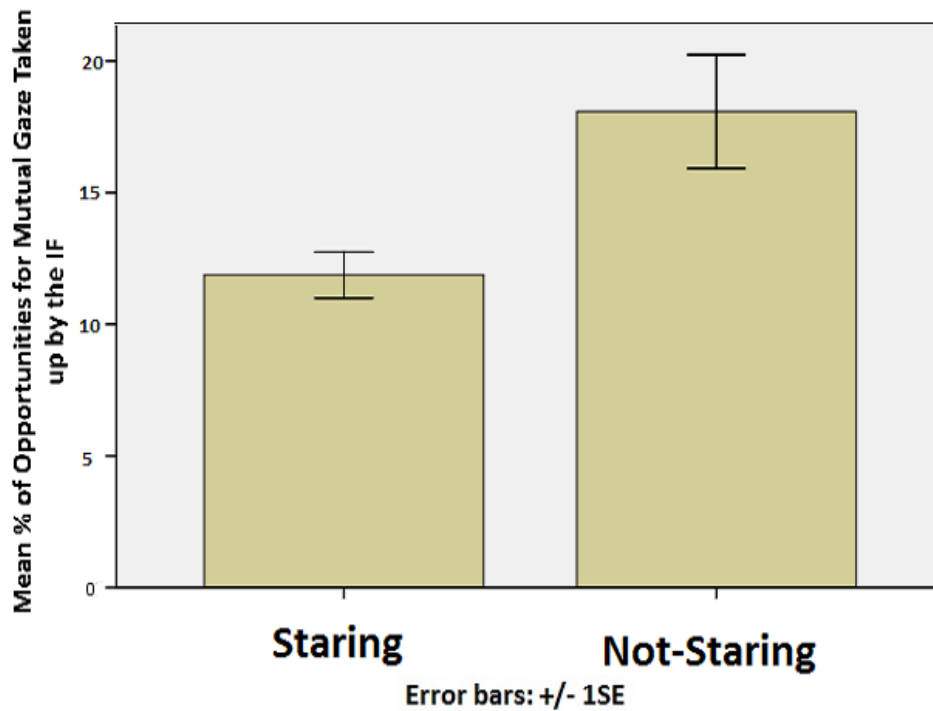


**Figure 3: Mean % of trial IF spent looking at IG's avatar.**

A paired samples t-test found that the IF spent significantly more time looking at the IG's avatar in the *not-staring* than in the *staring* condition,  $t(21)=2.705$ ;  $p<.05$  ( $M=14.96$ ,  $SD=5.81$  and  $M=11.87$ ,  $SD=4.12$ , respectively; see Figure 3).

To investigate the amount of mutual gaze that the dyad engaged in under each condition, the absolute amounts of mutual gaze were compared using a paired samples t-test. Although nearing significance, there was no overall difference found between the overall amount of mutual gaze in the 2 looking conditions ( $p>.05$ , ns).

To investigate this further, the proportion of the total number of opportunities for mutual gaze that were taken up by the IF was compared for the staring and not-staring conditions (see Figure 4).



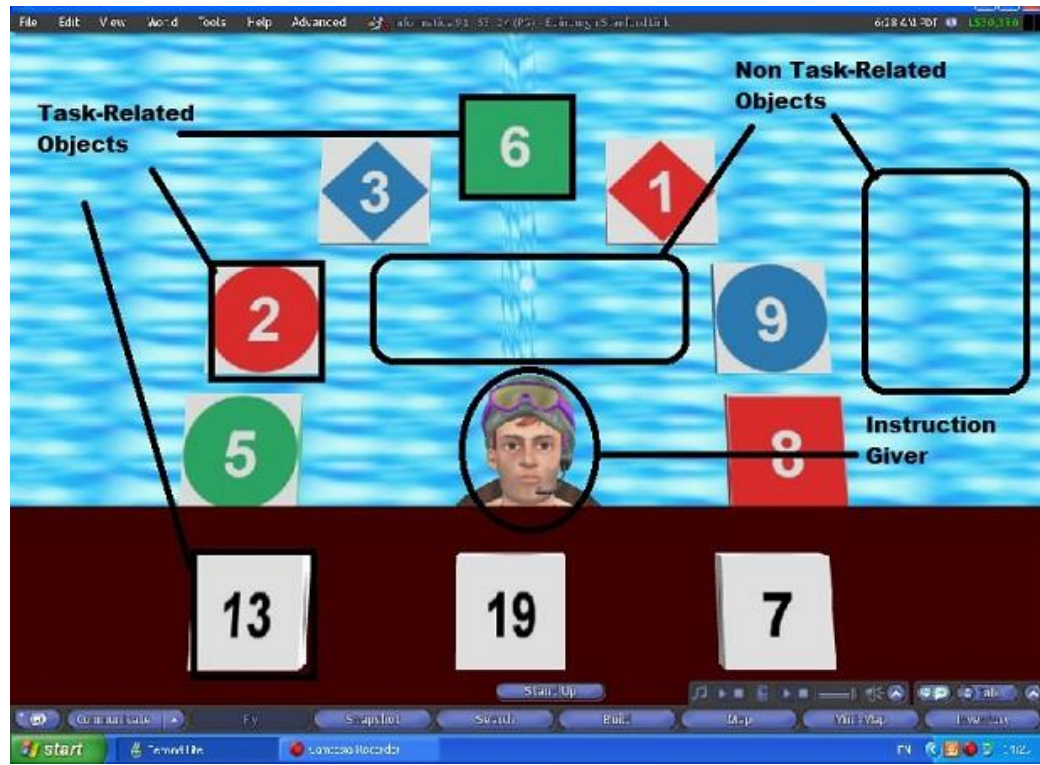
**Figure 4: Mean % of opportunities for mutual gaze taken up by the Instruction Follower**

The total opportunities for mutual gaze equated to all of the times when the IG was looking at the IF. When the IF looked back at the IG, these opportunities were said to be taken up. In the *staring* condition, this uptake was the same as % of the trial during which the IF looked at the IG (compare column 1 in Figure 3 and Figure 4). A paired-samples t-test found that there were significantly more opportunities for mutual gaze taken up in the not-staring condition than in the staring condition ( $M=18.08$ ,  $SD=4.12$ ;  $M=11.87$ ,  $SD=10.13$  respectively),  $t(21)=3.417$ ;  $p<.005$ .

### **3.2.2 Hypothesis 2 – gaze and task performance**

It was anticipated that there would be a positive correlation between the proportion of mutual gaze between the dyad and the task performance score in both the staring and the not-staring conditions: the more mutual gaze there is between the pair, the higher the task performance score. The task performance score was defined as number of correct answers out of a possible 15 in each block. It was further anticipated that there would be a negative relationship between task performance and the proportion of time that the follower spent looking at task-irrelevant objects. Task-irrelevant (or non-task related) objects are defined as anything within the screen

space that is not related to the completion of the task, i.e. anything apart from the instruction giver's avatar or the tiles (see Figure 5). Task-related objects were the instruction tiles, response tiles and IG's avatar.



**Figure 5: Areas of interest within the Instruction Follower's view**

Before analysis of the task performance scores, three of the dyads had to be excluded, since they had failed to understand the instructions, and therefore responded to the questions incorrectly. The remaining 18 dyads' task performance scores were compared.

A Wilcoxon Signed Ranks Test found there to be no significant difference between the overall task performance scores in the staring and not-staring conditions ( $Z = -.303, p = .71$ ). Indeed, median task performance scores were 14 in both conditions.

## Mutual gaze and task performance in the staring condition

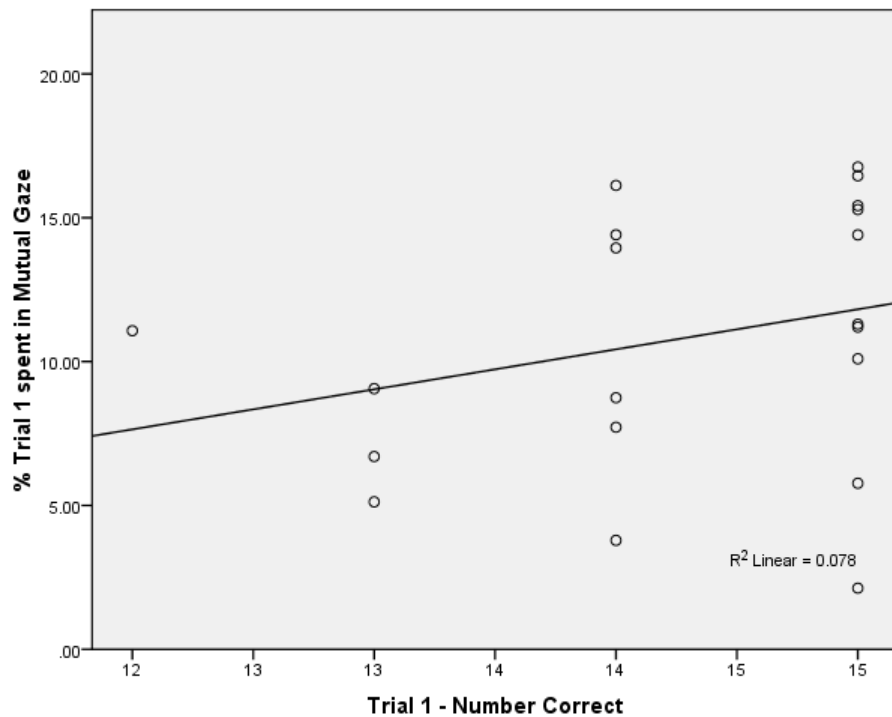


Figure 6: Mutual gaze and number correct in Staring condition

A Spearman's rho correlation found that in the *staring* condition, despite a positive trend, the relationship between task performance and mutual gaze failed to reach significance ( $r_s = .36$  (18),  $p(\text{one-tailed}) = .062$ ; see Figure 6).

## Mutual gaze and task performance in the not-staring condition

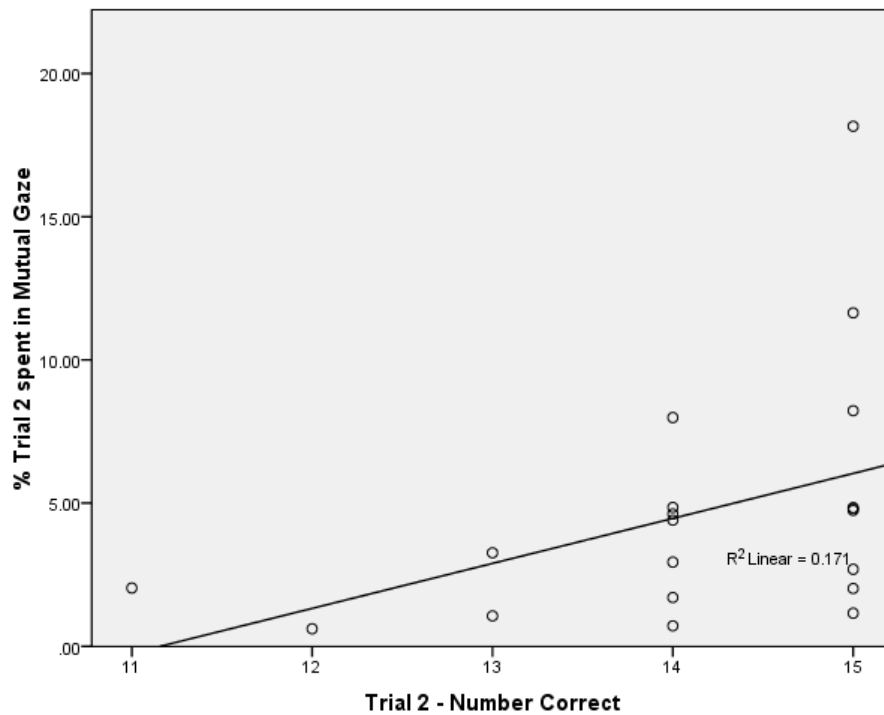
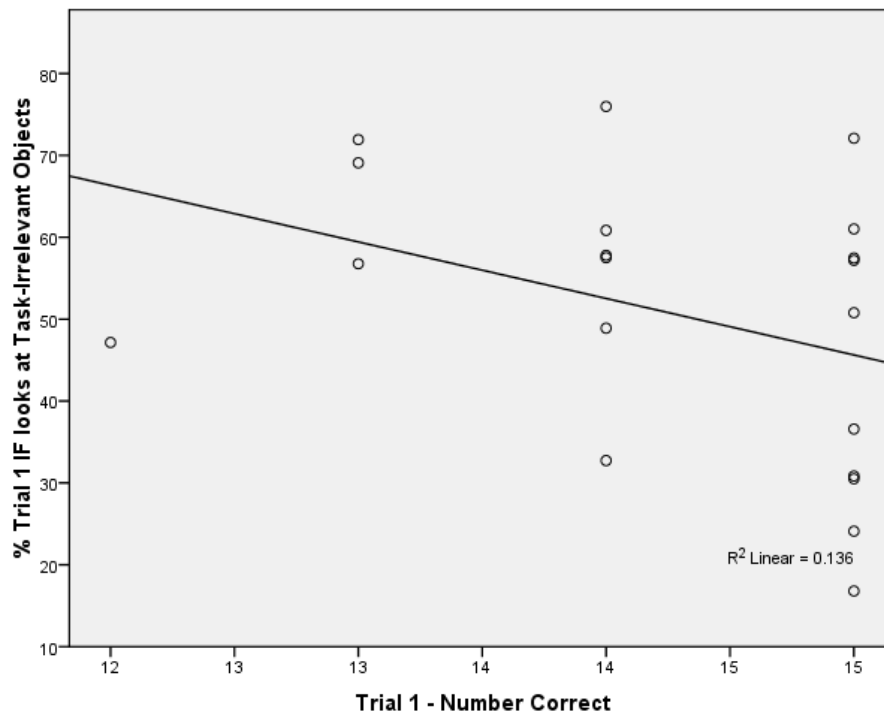


Figure 7: Mutual gaze and number correct in Not-Staring condition

A Spearman's rho correlation found that in the not-staring condition, task performance *was* significantly correlated with the proportion of trial spent in mutual gaze ( $r_s = .48$  (18),  $p(\text{one-tailed}) < .05$ ; see Figure 7).



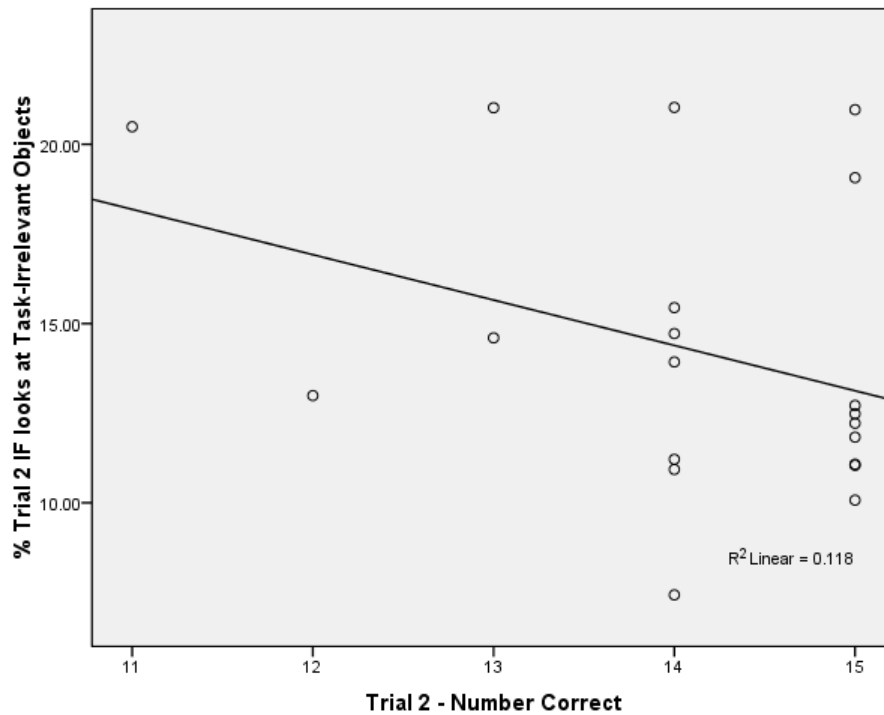
## Task Performance and IF Looking at Task-Irrelevant Objects in Staring Condition



**Figure 8: Number correct and % of trial spent looking at task-irrelevant objects in Staring condition**

In the *staring* condition, it was found that the negative relationship between task performance and the proportion of time that the follower looked at task irrelevant objects failed to reach significance ( $r_s = -.35$  (18),  $p(\text{one-tailed}) = .065$ ; see Figure 8).

## Task Performance and IF Looking at Task-Irrelevant Objects in Not-Staring Condition



**Figure 9: Number correct and % of trial spent looking at task-irrelevant objects in Not-Staring condition**

In the not-staring condition, it was again found that the negative relationship between task performance and the proportion of time that the follower looked at task irrelevant objects failed to reach significance ( $r_s = -.359$  (18),  $p(\text{one-tailed}) = .060$ ; see Figure 9).

### 3.2.3 Further investigation into the dyads' looking behaviour

It was of interest to investigate further the looking patterns of the dyads in the staring and not-staring conditions. The task-related and non-task related areas are described in Figure 5.

At any one time, the IG's avatar was looking at either the IF or the stimulus tiles. In the staring condition, there were 3 possible looking behaviours of the dyad:

1. Mutual gaze – the giver looks at the follower and the follower looks at the giver's avatar

2. Monitoring gaze – the IG looks at the follower, while the follower looks at a task-related area (a tile)
3. Other – the IG looks at the follower, while the follower looks at a task-irrelevant area – i.e. not the IG or a tile

In the not-staring condition, there were 6 possible looking behaviours of the dyad:

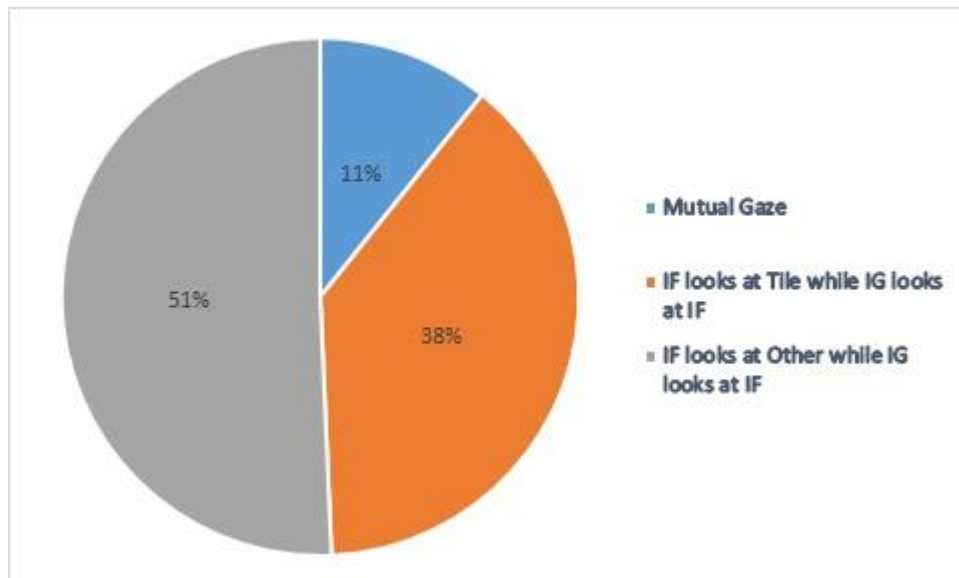
1. Mutual gaze – the giver looks at the follower and the follower looks at the giver's avatar
2. Monitoring gaze – the IG looks at the follower, while the follower looks at a task-related object (a tile)
3. Other – the IG looks at the follower, while the follower looks at a task-irrelevant object – i.e., not the giver or a tile
4. Monitoring gaze – the giver looks at a task-related object, while the follower looks at the giver
5. Shared gaze – both the giver and the follower look at a task-related object
6. Other – both giver and follower look at a task-irrelevant object

Since there was no overall difference between the task performance scores in the staring and not-staring conditions, it was of interest to investigate why mutual gaze had a differing effect on task performance in the two conditions. The IF's eye movements during missed opportunities for mutual gaze were compared under the two independent variables. This comprised all of the occasions when IF did not look at IG in the staring condition, compared with all the times in the not-staring condition when IG is looking at IF, but IF is looking elsewhere (i.e. not engaging in mutual gaze). The distribution of IF looking behaviour during all such opportunities for the staring and not-staring conditions, can be seen in Figure 10 and Figure 11, respectively. In the not-staring condition, this time comprised approximately 27% of the total trial.

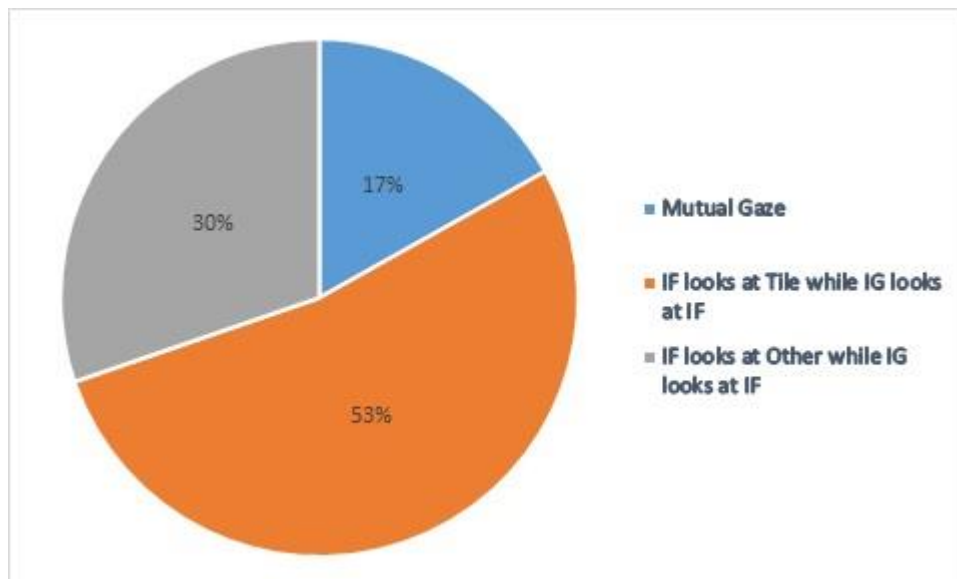
The question of interest is: where does the follower tend to look when not looking at the IG, and is this pattern of looking different in the staring and not-staring conditions? In the staring condition, the follower's gaze should be deflected from the giver (social reasons will make the follower avert their gaze), but in the not-staring condition, any looks away from the giver will be directed towards task-

relevant objects. It is anticipated, therefore, that followers in the staring condition will avert their gaze to task-irrelevant objects, rather than task-related tiles, since this deflection of gaze will be towards ‘anything that is not the giver’, compared to, in the not-staring condition, towards a task-relevant stimulus (i.e. a stimulus or response tile). It is hypothesised, therefore, that there will be a difference in the amount of looking at ‘other’ and at ‘task-related’ / ‘tile’ in the staring and not-staring conditions.

Figure 10 and Figure 11 show the looking behaviour of the instruction follower during the time in which the instruction giver is looking at the follower. This constitutes 100% of the time in the staring condition, and approximately 27% of the time in the not-staring condition. They show that that when the IG is looking at the IF, the follower spends proportionally more of the trial looking at non-task-related areas, or ‘other’, in the staring condition than in the not-staring condition. A paired t-test found this difference to be significant. The proportion of the trial that the IF spends looking at non-task-related areas compared with task-related areas, was found to be significantly higher in the staring condition than in the not-staring condition ( $t(19)=3.509$ ;  $p<.01$ ).



**Figure 10: Mean % of trial spent in each looking behaviour in the Staring condition**



**Figure 11: Mean % of trial spent in each looking behaviour in the Not-Staring condition. This only includes times during which there were opportunities for mutual gaze.**

### **3.2.4 Summary**

Staring did, indeed, show evidence of suppressing mutual gaze, as demonstrated by a lower uptake in the staring condition. As predicted, mutual gaze facilitated task performance, but only in the not staring condition; there was no such relationship when the follower was being stared at. It was found that when being stared at, users are more likely to look at task-irrelevant (rather than task-related) objects; they are more likely to attend to task-related objects when not being stared at. These results will be discussed in chapter 4.

### **3.3 Discussion**

This discussion is based, in part, on a paper presented at The 33rd Annual Conference of the Cognitive Science (Dalziel-Job, Oberlander, & Smith, 2011). Experiment 1 aimed to examine the function and patterns of gaze and mutual gaze during computer-mediated task-oriented interactions. It also aimed to further develop a novel paradigm that enables research into users' eye movements during interaction with agents/avatars in 3D virtual environments.

#### **3.3.1 Staring and mutual gaze**

We were interested here in the impact of constant staring by one conversational partner at another on mutual gaze between the dyad. It was found that if an Instruction Follower is being stared at, he is likely to spend less time looking at the face of the person staring – the Instruction Giver. Initial findings showed that staring did not affect the absolute amount of mutual gaze between the dyad: there was no significant difference between the total amounts of mutual gaze in the staring and not-staring conditions. The IF, however, had the opportunity to engage in mutual gaze at any time during the interaction in the staring condition, but there were fewer opportunities for mutual gaze in the other condition (approximately 27% of the not-staring, vs. 100% of the staring trial). Despite the increased opportunities, there were no more overall occurrences of mutual gaze in the staring condition than in the not-staring condition. In fact, a higher proportion of opportunities for mutual gaze were taken up in the not-staring condition than in the staring condition. This suggests that staring does have the potential to reduce mutual gaze uptake, which would imply that incorporating such looking behaviour into an experimental paradigm may not be ideal for this area of eye movement research.

It seems entirely reasonable to assume that there are social factors at work here, which discourage an individual from returning the stare of their conversational partner, to avoid being, as Kendon suggests, “vulnerable to him”. It could be argued, however, that the IF looked more at the IG during the not-staring condition because of visual information that could assist in the completion of the task in this condition.

Although this information is strictly redundant, this possible explanation will be tested in the next study with an additional baseline condition where the IG still looks at the stimulus tiles while describing them (redundantly), but does not look at the IF during the procedure. Comparison between the conditions will help distinguish attention attracted for task-related reasons (i.e. because the IG is looking at the tiles) from that attracted for social reasons (i.e. because something else is compelling the IF to look at the IG and engage in eye contact).

### **3.3.2 Mutual gaze and task performance**

The tasks were found to be easy, with many participants completing them all correctly. No difference in task performance was found between the staring and not-staring conditions (the median score for both conditions was 14/15). This suggests there could have been a ceiling effect, which may have masked some differences. Despite this, some significant task performance outcomes were found, unlike in previous similar studies (Blascovich et al., 2002; Zambaka et al., 2007). A further study would include tasks with higher levels of difficulty, interspersed with tasks with lower levels of difficulty. This would serve two purposes: firstly it could remove ceiling effects, and secondly, it could tease apart social facilitation / inhibition effects as they relate to agent/avatar interactions. Reaction times would also be recorded to produce a fuller picture of task performance.

As predicted, the more mutual gaze there was between a dyad, the better the task performance. This only held true, however, when the IF was not being stared at (although there was a positive trend in the staring condition). This suggests that if you want your interlocutor to retain the information that you are imparting, then you should try and maximise the amount of mutual gaze between the pair of you. But this does not involve staring: staring will not influence task performance in the same way that not-staring can; staring maximises neither mutual gaze nor task performance. In the next study we will systematically explore the relationship between varying amounts of gaze by the IG and its effects on mutual gaze and task performance.

### **3.3.3 Further investigation into the dyads' looking behaviour**

The finding that the IFs were less likely to spend their non-mutual-gaze periods looking at task-related objects in the staring condition than in the not-staring condition may go some way towards explaining the lack of relationship between mutual gaze and task performance when being stared at. Directing gaze towards irrelevant objects does not help task performance. Gaze aversion in the staring condition would seem to be driven by social factors (discomfort at being stared at), looks away from Tracker whereas in the not-staring condition were task-related (looking away in order to look at task-related objects and gain information about the task).

### **3.3.4 Next steps**

In the current study, the participants were fully aware that they were interacting with another human, the avatar behaviour was human-like (it imitated human behaviour, although how human-like was not quantified) and there is precedent for using virtual humans to investigate human-human interaction (Yee et al., 2007; Bailenson et al., 2001). But at this point, strong conclusions about face-to-face, human-human behaviour cannot be drawn.

Chapter 4 introduces the second study in the program. In this human-agent paradigm, the dependent variable agency will be included, meaning that users will either be told they are interacting with an avatar (human controlled) or an agent (computer controlled). This should foreground the differences between how people treat humans and computers within this paradigm.

The human-agent experiment will aim to establish how much looking by one conversational partner at another is optimal for mutual gaze and task performance on a given task. If mutual gaze can be maximised, then it follows that task performance may also be optimised. The experiment covered in chapter 4 takes the form of a human-agent experiment within Second Life, using a similar paradigm. Additional control conditions will be in place to help eliminate other possible explanations for variation in looking behaviours.



## 4 Chapter 4: Experiment 2 – Human-Agent Experiment

Experiment 1 involved a task-oriented interaction between 2 participants via avatars in Second Life. The assumption that having one person stare continuously at the other is an appropriate method for mutual gaze research has been successfully challenged; instead of providing maximum opportunities for mutual gaze, which were then taken up (as was presumably the intention), incorporating such looking behaviour into a paradigm is likely to introduce experimental artefacts, resulting in reduction of mutual gaze. Given the aforementioned benefits of mutual gaze, the current study aims to extend the human-human results by establishing: What *is* the optimum amount of looking by one conversational partner at another for maximising mutual gaze between the dyad? Further to this, social presence measures (manipulated and perceived) are included in the paradigm to investigate the relationship between gaze, mutual gaze, task performance and social presence.

This study aimed to uncover participants' social perceptions of an interlocutor during a task-based interaction. Furthermore, it aimed to establish the optimum amount of looking by one conversational partner at another in order to maximise the amount of mutual gaze between the dyad, task performance scores and perceptions of social presence. It was anticipated that if Tracker had social influence over the users, then we would find social facilitation effects on the task performance (as found in experiment 1). This is assuming the tasks are well-learned, which arithmetic can be said to be, especially among a sample of predominantly university students. Behavioural measures – eye movements or task performance – are likely to be able to better detect differences in social presence than a self-report questionnaire. We would, therefore, expect to find more effects in the eye movement and task performance data than in the questionnaire data. We would also expect social presence ratings to vary as a function of looking behaviour (as found in Bailenson, Blascovich, Beall, & Loomis, 2003), and for participants in the avatar condition to perceive their interlocutor to be more socially present than those in the agent condition.

From a methodological point of view, the human-agent study was designed to address the critique that the extra looking at Tracker in the not-staring condition in experiment 1 was not an artefact of the extra task-related visual attention (i.e. looking at the tiles while describing them).

Experiment 2 was run, on the whole, within the same paradigm as experiment 1, using the same stimuli and set-up. The main difference in the design of the study was that instead of 2 people interacting, there was only 1 participant who was interacting with a programmed avatar within Second Life. The programmed avatar will be referred to as either Tracker, or the Instruction Giver (IG) or giver throughout. The participant will either be referred to as such, or as the Instruction Follower (IF) or follower.

#### ***4.1 Research questions – Gaze, task performance and social presence***

This study investigates the relationship between mutual gaze, perceived and manipulated social presence and task performance during a task-focused interaction within Second Life? The research questions are as follows:

1. What is the optimum amount of looking by one individual at another for maximising:
  - Mutual gaze
  - Task performance
  - Perceived social presence?
2. What is the relationship between purported agency (manipulated social presence) and mutual gaze, task performance and perceived social presence? Can behavioural measures in this paradigm – gaze/task performance – identify variations in actual social presence which go undetected by a self-report questionnaire?
3. What is the function of gaze, specifically mutual gaze, during task-based interactions within Second Life?

The following section reports on the methods, adapted from the human-human paradigm, that were used to investigate these questions.

## 4.2 Method

### 4.2.1 Independent Variables

**Table 1: Human-Agent experiment independent variables**

Agency (between participants)	Instruction Giver Looking Condition (within participants)			
	0% looking	25% looking	50% looking	75% looking
Agent	0% looking	25% looking	50% looking	75% looking
Avatar	0% looking	25% looking	50% looking	75% looking

#### 4.2.1.1 Agency

In this mixed (4x2 within-between) design experiment, the first independent variable was the purported agency (agency) of Tracker, the user’s conversational partner and Instruction Giver throughout the tasks. All participants interacted with a programmed avatar within Second Life. Prior to the interaction, all participants received a written set of instructions. Half of the participants’ instructions included the information that “You will be interacting with Tracker, a computer-controlled agent – an Instruction Giver – in Second Life, an online virtual environment, in order to complete arithmetic problems” (see Appendix C). This was the Agent condition. The other half of the participants’ instructions (in the Avatar condition) included “You will be interacting with another person – an Instruction Giver – through Tracker, an avatar in Second Life, an online virtual environment, in order to complete arithmetic problems” (see Appendix D). Other than this, instructions were identical for both groups, as were all other manipulations. This produced the condition Agency, a manipulation that was designed to lead participants to believe that they were either interacting with a human or with a computer-controlled agent, thus potentially manipulating the participants’ social perceptions of their interlocutor. This was a between-participants manipulation.

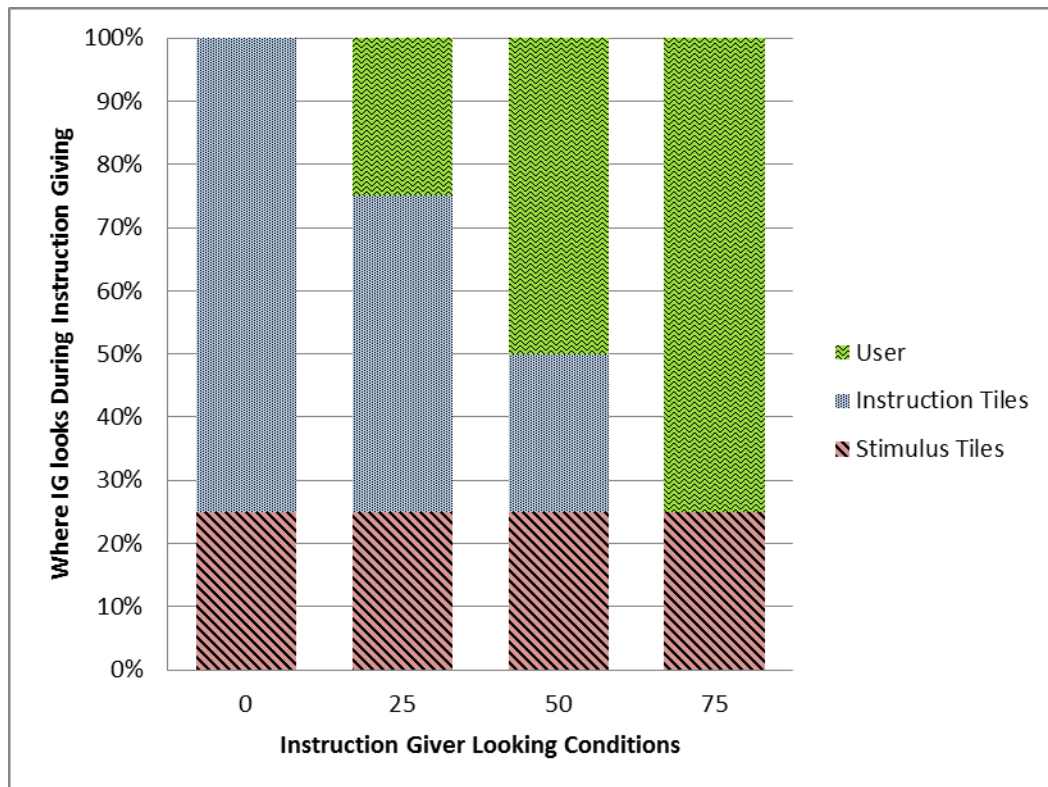
#### 4.2.1.2 Looking Behaviour

The second independent variable was the *looking behaviour* of Tracker, the programmed avatar. One of the research questions was to establish the optimum amount of looking by one conversational partner at another, in order to maximise mutual gaze, social perceptions and task performance. To attempt to answer this question, Tracker was programmed to look at the user during 0%, 25%, 50% or 75% of the time during which she was giving instructions (see Figure 12).

A critique of the previous study was that in the not-staring condition, the Instruction Giver's avatar looks at the tiles as he describes them, compared with the staring condition, where the IG's avatar stares at the Instruction Follower and does not look at the instruction tiles at all. As previously discussed, the looking in the not-staring condition was redundant looking: the IF was receiving verbal instructions as well, so was not relying on the looking of the IG in order to complete the tasks. There should, therefore, have been no task-related bias in the not-staring condition. In order to completely rule this possibility out, however, in the current study the IG looked at the instruction tiles while describing them in all of the looking conditions. This eradicated any bias due to additional visual instruction being provided, as per the critique of the not-staring condition in the previous study. Tracker, therefore, looked at the stimulus tiles as she described them in all of the looking conditions (this took up 25% of the instruction-giving time – approximately the length of time taken to describe each tile). This means that, along with the verbal instructions, the user was provided with the same amount of visual information to assist in the completion of the task during each looking condition. During the remaining 75% of the instruction-giving time, Tracker looked at the user either not at all (giving the 0% condition), a third of the time (25%), two thirds of the time (50%) or all of the time (75%).

During any time not spent looking at stimulus tiles or at the user, Tracker looked down at her instruction tiles. These were out of sight of the participant. The 25% condition approximated the looking behaviour in the not-staring condition of experiment 1; in that experiment, the IG's natural behaviour was to move his avatar's gaze towards the user during approximately 27% of each trial (the avatar looked at the instruction tiles for the remaining 75% of the trial). The looking behaviour of the

IG in that experiment was driven by the instruction giver so could be expected to approximate to natural gaze patterns in that – and the current – paradigm (taking into account the fact that the avatar was also required to look at the instruction tiles). The 75% condition equates to Argyle’s (2013) assertion that listeners gaze at a speaker during approximately 75% of a dyadic interaction. The 0% condition provided a baseline and the 50% condition was included to provide a reference point between 25% and 75% looking. It was anticipated, based on the human-human experiment and Argyle (2013) that either 25% or 75% looking by Tracker would produce the most naturalistic looking behaviour.



**Figure 12: Tracker’s looking behaviour during experiment 2**

Looking behaviour was a within-participants variable; all participants experienced all 4 looking behaviours. These behaviours were counterbalanced between participants to avoid any undesired learning effects due to order of presentation. The tasks were systematically varied across conditions to avoid any erroneous effects due to the items themselves. The order of looking conditions encountered was the same in each agency condition; for each participant in the Agent condition there was a

corresponding participant encountering an identical order of conditions and tasks in the Avatar condition. The only difference, therefore, between the Agent and Avatar conditions was the instructions administered at the start of the experiment.

## **4.2.2 Dependent Variables**

### **4.2.2.1 Gaze**

It was of interest to discover which Tracker looking behaviour resulted in the most mutual gaze; does more giver-looking at the follower result in more mutual gaze between the dyad? We have previously discovered that staring does not maximise the amount of mutual gaze, but what about if the giver looks at the follower 25% of the time, or 50%, or 75%? There will clearly be no mutual gaze in the 0% condition; this level was set to act as a baseline measure – how much attention do the instruction giver's eyes attract when she doesn't look at the follower at all?

The first dependent variable was proportion of opportunities for mutual gaze taken up by the participant – the instruction follower. The absolute amount of mutual gaze was also measured as a proportion of the total trial time. Finally, the proportion of the trial that the participant looked at the instruction giver's eyes, regardless of whether the giver was looking at him, was also measured.

### **4.2.2.2 Task Performance**

Maximising mutual gaze between the dyad was not the only purpose of manipulating the instruction giver's looking behaviour. It was also of interest to establish if any of the conditions resulted in maximised task performance. Furthermore, would there be a positive relationship between mutual gaze and task performance, as previously discovered, or would an increased amount of mutual gaze result in lower task performance or have no effect? Two measures of task performance were utilised.

#### ***Number Correct***

The first task performance measure was proportion of tasks completed correctly, as found in experiment 1. Each participant encountered 10 tasks in each of the 4 looking condition, resulting in 4 separate scores out of 10 for each individual, ranging from 0/10 to 10/10.

### ***Reaction Time***

Additionally, reaction times were recorded. Initially, two such measures were taken:

1. The time from the start of each trial (i.e. from when the stimuli changed and the onset of verbal instructions for each task) until a response was selected by the IF (*RT Start*) and
2. The time taken from when the instructions ended for the IF to select a response (*RT End*).

Since the instructions were recorded, the amount of time taken for the instructions to play was the same for each task across conditions. It is worth noting here that the 2<sup>nd</sup> measure (*RT End*) could be recorded as 0, since a response could be selected before the instructions were finished.

The written instructions at the start encouraged participants to complete the tasks as quickly as possible, and they were further incentivised by being told that they had the potential to win an Amazon voucher if they were one of the 5 people to complete the tasks in the fastest overall time.

### **4.2.2.3 Social Presence Measures**

The 3<sup>rd</sup> measure was the perceived social presence of the instruction giver. Since existing social presence questionnaires had been designed and tested for use within a specific application, and therefore validity for use within this forum could not be guaranteed, a new questionnaire was designed by selecting questions from an existing presence questionnaire and adapting them to capture aspects of the interaction related to social presence, as well as presence as a whole (Witmer & Singer, 1998). The resulting questionnaire was made up of 20 items, and the

participants were presented with 5 possible responses: ‘Not at all’, ‘Slightly’, ‘Moderately’, ‘Fairly’, ‘Extremely’. The results of the questionnaire were subjected to a Principal Components Analysis in order to extract factors for further investigation (see Appendix E for all items making up the questionnaire). These questions were administered 4 times to each participant, once after each of the 4 blocks of tasks (i.e. after encountering each looking condition). They were presented to the participant, and responded to within Second Life. Each set of questions, therefore, related to the follower’s experience under a different looking condition. Tracker was not present while the questionnaire was being completed.

#### **4.2.3 Participants**

48 participants (mean age 22.2; 21 male) were recruited via SAGE (Edinburgh University’s employment website) and by word of mouth. They were asked not to apply if they had participated in the previous study. Each participant was a volunteer and received £6.50 financial remuneration for participating in the study. Sessions lasted between 45 and 60 minutes. Each participant passed an Ishihara Test for colour-blindness (Ishihara & Force, 1943). Of the original participants, the data from 5 had to be discarded and additional participants recruited to fill those conditions. This was due to issues ranging from Second Life difficulties (bugs / crashes) and software issues (problems with sound/video recording) to hardware failure. The final number of usable participants was 43. To preserve the pairing of participants from each agency condition, the results for one of the participants were discarded, leaving usable data for 21 pairs of participants.

#### **4.2.4 Apparatus**

Participants viewed the experiment on a 19 inch CRT display, sitting approximately 40cm away from it. They were free to interact with a standard mouse, but asked not to touch the keyboard. Participants were linked up to an SR-Research EyeLink II head-mounted tracking system. They were able to hear the verbal instructions from Tracker via speakers attached to the PC. The volume at which the instructions were heard could be adjusted to suit the instruction-follower. The procedure took place within a sound-proofed studio. The sample rate was set at 500Hz and the participants’ dominant eye was tracked monocularly. A 9-point calibration matrix



was used at the start of each participant's experiment and between each block / looking condition. Camtasia Studio (produced by TechSmith) recorded what each participant could see on the screen throughout the procedure. Movie files were generated, which incorporated what the instruction-follower could hear throughout the experiment with what they could see to allow for easier analysis. Finally, the program controlling Tracker produced a log file after each trial. Each of these files recorded and time-stamped everything that happened during that trial, including stimulus changes, button presses and onset of audio instructions/videos.

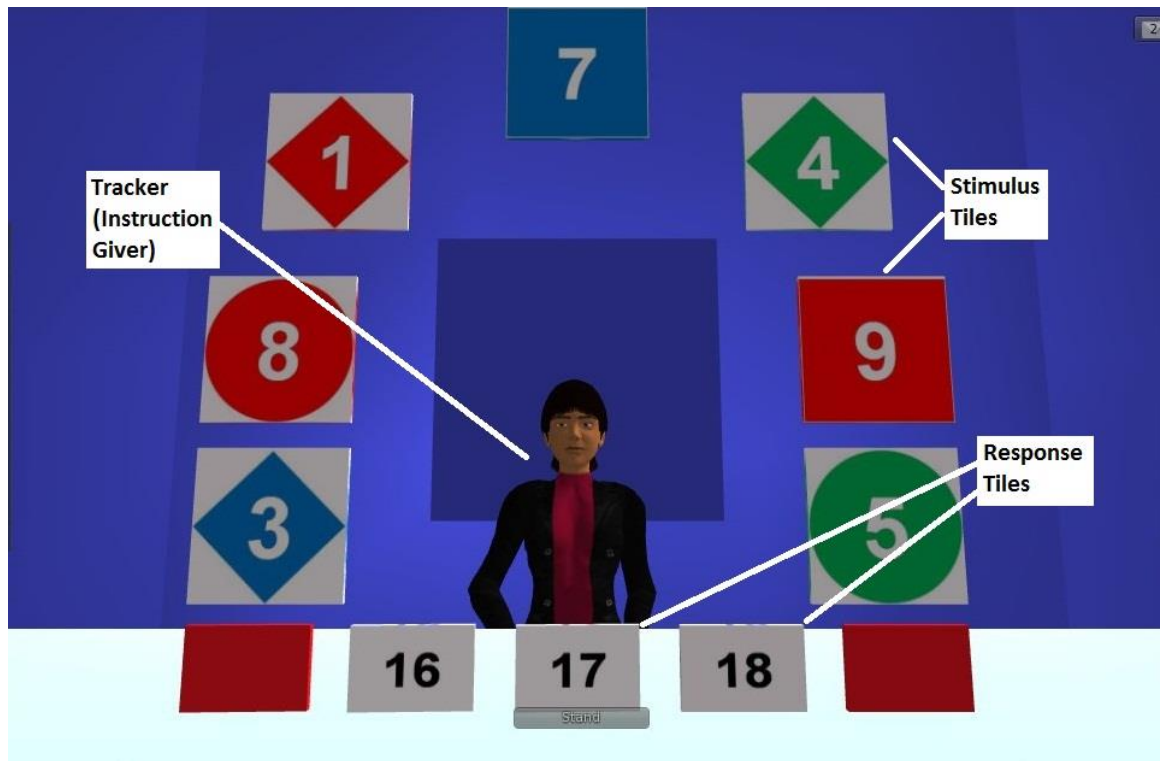
#### **4.2.5 Design**

In a mixed design, all participants carried out 10 tasks under each of the 4 looking conditions: 0%, 25%, 50%, and 75%. The order of these conditions and the tasks themselves were counterbalanced between the participants for the 4 conditions to control for effects due to learning and due to the tasks themselves. Tasks under each condition were presented in separate blocks, with re-calibration between each block. Each participant was either told that they were interacting with an avatar (another person) or an agent (a computer) before the start of the experiment. Participants were randomly assigned to one of the two agency conditions.

#### **4.2.6 Stimuli**

A building comprising 1 large closed room was built on VUE, the Edinburgh University Island within Second Life (see <http://vue.ed.ac.uk/>). Within the room were 2 chairs facing each other with a glass screen between them. One of these chairs was for the participant's avatar and the other for Tracker, the Instruction Giver. In this study, due to the availability of suitable voices, Tracker appeared as a female (she was a male avatar in the human-human study), although her appearance was fairly androgynous. Between the two chairs was a glass panel, upon which the stimulus tiles were projected. The participant's view was such that they could see Tracker giving the instructions, as well as the 7 stimulus tiles and 3 response tiles, from which they were to select one by clicking on it to register their answer to each task (see Figure 13). Each stimulus tile contained a number surrounded by a shape (a diamond, a circle or a square) in one of 3 colours (red, green or blue). There was no

more than 1 stimulus tile with each shape-colour combination visible at any one time. Each response tile contained a black number on a white background.



**Figure 13: Participant's view of Tracker**

The participant's chair within the virtual environment was programmed so that when an avatar sat upon it, the camera view – i.e. what the participant controlling the avatar could see – was taken over. The camera view for the IF's avatar was such that each participant saw through their avatar's eyes. This was designed to ensure that the participant had the optimum view of Tracker and where she was looking, as well as the stimulus and response tiles.

Each block consisted of 10 tasks. Each task was a short, relatively simple, arithmetic problem consisting of 3 numbers to be added, subtracted or multiplied. There were 5 easier and 5 more difficult tasks presented and the order was easy-hard-easy-hard-etc. The easier ones involved adding and subtraction only, and the harder also included some multiplication. This was to try and avoid the ceiling effect found in the human-human study, without making the tasks so challenging that they would

result in a floor effect. The order of the easy/difficult tasks was the same for all participants. The tasks were piloted outside Second Life before running the experiment and were judged to be neither very hard nor very easy.

Each task was presented by Tracker to the participant verbally. Tracker relayed the numbers involved by describing the colour and shape on the stimulus tile (as the IG did in the previous study). An example instruction from the stimuli in Figure 13 would be “green diamond, multiplied by green circle, plus blue square, minus red square” and the correct response tile to click on would be 18. This was one of the more difficult tasks, including a multiplication. Upon the IF clicking on one of the response tiles, all of the texture on the tiles would change for the next task.

The questionnaire was textures were projected in a screen behind Tracker’s seat. The 20 questions included items such as “How much were you able to control Tracker’s behaviour?” and “How much did Tracker’s verbal instructions involve you in the interaction?” The responses, in the form of a 5-point Likert scale, were “Not at all”, “Slightly”, “Fairly”, “Moderately” and “Extremely”. These were projected as textures onto the response tiles in front of the participant’s avatar. All questionnaire and response textures appeared in black text on a white background.

All of the tiles – the instruction, stimulus, questionnaire and response tiles - were created within SL and each of the stimuli was imported into SL and attached as a texture to the tile when needed for the task. This was controlled by a program which also controlled Tracker’s behaviour within Second Life. All stimuli – 172 in total – were created in Microsoft PowerPoint and edited using Microsoft Paint version 5.1 and the GIMP (GNU Image Manipulation Program).

#### **4.2.7 Procedure**

Participants arrived at the Joint Eye-tracking Laboratory at Edinburgh University. Each was seated in front of a computer and asked to read and sign a consent form outlining the experiment (see Appendices F and G). They were then given a short set of written instructions which gave them a brief overview of what the experiment consisted of. The instructions for the agent condition informed them that they were interacting with an agent in Second Life, which was controlled by a computer (see Appendix C). In the avatar condition the instructions told them that they were

interacting with another human via an avatar in Second Life (see Appendix D). Other than this the instructions were identical for both conditions. The experimenter checked that the participant understood the instructions before starting the experiment.

Camtasia screen recorder was initialised. It was then determined which was the subject's dominant eye, which would consequently be tracked, and the eye tracker was placed onto the participant's head. It was explained that, while the head set with the eye tracker had to be secure enough not to move, it must still be comfortable, and that it would be tightened until the participant felt that it was tight enough. Following this, Second Life was started up and the IF's avatar was signed in. Once SL, Camtasia and the hardware had been set up, the participant was led through the calibration process. After this, the program controlling Tracker and the texture-changes on the tiles was initialised for one of 4 blocks of tasks. Each of these blocks contained 10 tasks – 40 in all – and the sets of tasks were counterbalanced between all of the experimental runs to control for possible artefacts due to materials.

After the participant had completed one block of tasks, Tracker disappeared and the questionnaire questions appeared as textures on the dark blue screen behind where Tracker had been sitting (see Figure 13). The questionnaire responses were presented as textures on the response tiles. Once the 20 questions had been answered, the participant was re-calibrated and block 2 was started. This happened for each of the 4 blocks of tasks and questionnaires. The same questionnaire items were presented in the same order after each block. Upon completion of all four trials and the associated questionnaires the participants were asked to sign a receipt and were given £6.50. After this, the purpose of the study was disclosed, as well as a brief tour of SL if requested.

### **4.3 Results**

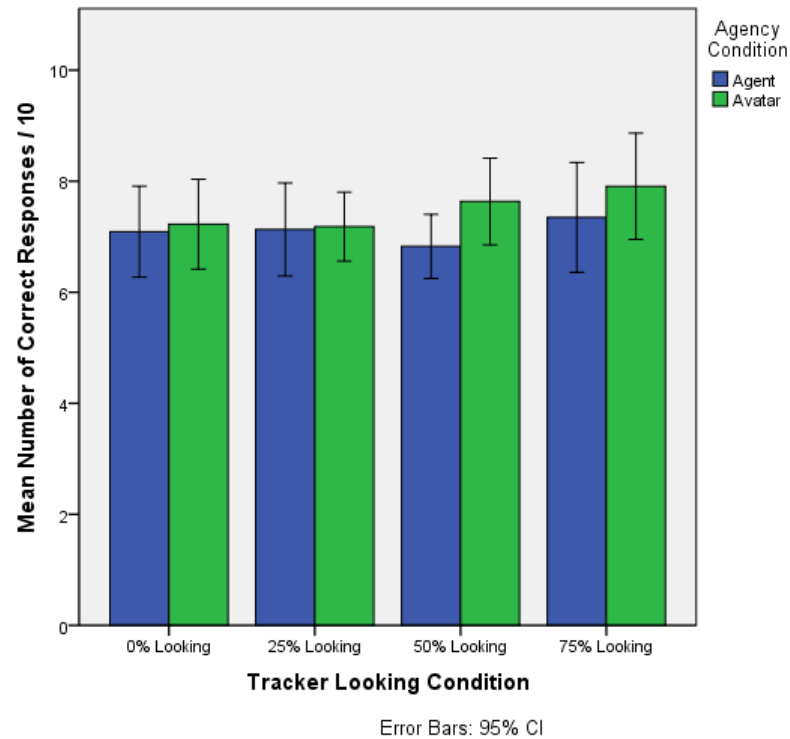
To expand on the human-human experiment results, the human-agent experiment aimed to discover what the optimum amount of looking by one conversational partner at another was, in terms of mutual gaze, task performance and social

perceptions of the instruction giver. It was anticipated that different amounts of looking at user by the IG (0%, 25%, 50% and 75%), as well as which agency condition the user was assigned to, would result in variations in looking behaviour, task performance and perceptions of the interlocutor and the interaction. Task performance (reaction time and number of tasks correctly completed) was initially examined, followed by the user's gaze (fixation duration, mutual gaze uptake, interest area analysis). The relationship between mutual gaze and task performance was then investigated. Next the questionnaire data was subjected to a principle components analysis and the resulting factor scores were analysed. Correlations were carried out between factor scores and both mutual gaze and task performance. Finally, a logistic regression was performed with factor score for each factor, mutual gaze and reaction time as predictors and agency as the binary outcome variable.

### **4.3.1 Task performance**

#### **4.3.1.1 Number Correct**

The first task performance measure was proportion of tasks completed correctly. Each participant encountered 10 tasks in each of the 4 looking conditions, resulting in 4 separate scores out of 10 for each individual, ranging from 0/10 to 10/10.

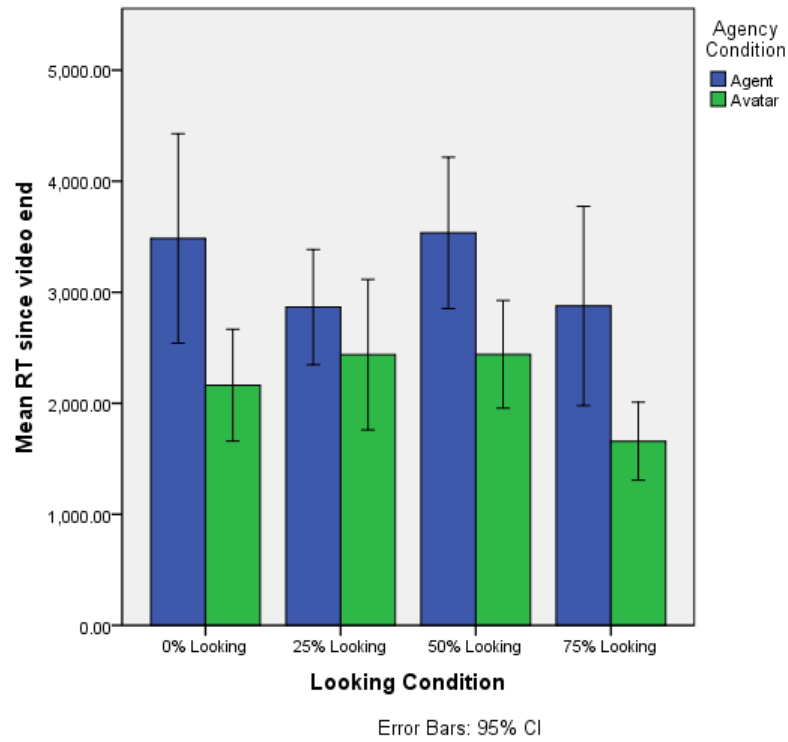


**Figure 14: Mean number of correct responses by Agency and Looking Condition**

No looking condition  $\times$  agency interaction was found for number of correct responses out of 10 ( $F(3) = .665$ ,  $p = .575$ ; see Figure 14). The main effect of looking condition on number of correct responses out of 10 was also found not to be significant ( $F(3) = .435$ ;  $p = .73$ ). There was also found to be no significant effect of agency on number of correct responses ( $F(1) = 1.931$ ;  $p = .172$ ).

#### **4.3.1.2 RT from end of Instructions (RTEnd)**

The time in ms that elapsed between the end of each set of instructions and the IF selecting a response was investigated. This was occasionally recorded as 0 if the IF selected a response before the conclusion of the instructions.



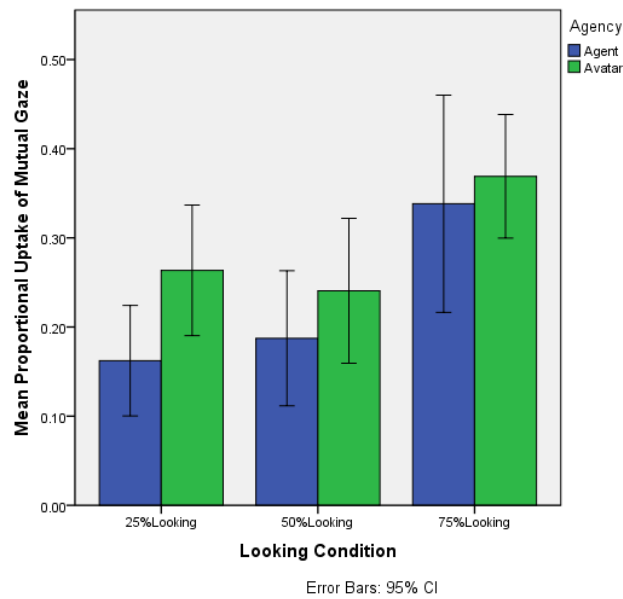
**Figure 15: Mean RT since end of instructions by Agency and Looking Condition**

No looking condition  $\times$  agency interaction was found for reaction time since video end (i.e. from the end of the instructions) ( $F(3, 126) = .956$ ;  $p = .416$ ; see Figure 15). There was, however a main effect of agency on RT end, with tasks being completed significantly faster in the Avatar condition than in the Agent condition ( $F(1, 42) = 13.705$ ;  $p < .01$ ;  $\eta_p^2 = .246$ ; Means: 2175, 3190; sd: 1191, 1757, respectively). There was no main effect of looking condition on RT end ( $F(3, 126) = 2.525$ ;  $p = .085$ ;  $\eta_p^2 = .051$ ). When looking at the Agent condition on its own, there was no significant difference between the RT end in the 50% and the 75% looking conditions ( $t(21) = 1.478$ ;  $p = .154$ ; Means: 3535, 2876; sd: 1537, 2024, respectively). In the Avatar condition, however, the RT since video end was found to be significantly shorter in the 75% looking condition than in the 50% looking condition ( $t(21) = 3.387$ ;  $p < .001$ ;  $d = 0.82$ ; Means: 1658, 2988; sd: 2024, 1537, respectively). Although, within the avatar condition, mean RT was shorter in the 75% looking condition than either 25% or 0%, these differences were not found to be significant (a Bonferroni adjustment was made for multiple comparisons).

## 4.3.2 Gaze

### 4.3.2.1 Mutual Gaze Uptake

The first gaze DV to be examined was mutual gaze uptake. This was calculated as the proportion of the total number of opportunities for mutual gaze (the time when IG was looking at the IF) that were taken up by the participant and which resulted in mutual gaze. This DV may be referred to as uptake of mutual gaze or simply mutual gaze (since absolute amount of mutual gaze is not analysed here).



**Figure 16: Mean proportion of mutual gaze uptake by Agency and Looking Condition**

No agency x looking condition interaction on mutual gaze uptake was found ( $F(2,64) = .797$ ;  $p = .455$ ; see Figure 16). There was a significant main effect of Looking Condition on the mean proportion of mutual gaze uptake ( $F(2,64) = 14.862$ ;  $p < .001$ ;  $\eta_p^2 = .317$ ). A paired samples  $t$ -test revealed a significant difference between the 25% and 75% looking conditions ( $t(33) = 4.975$ ;  $p < .001$ ;  $d = -0.91$  Means: .21, .36; SD: .15, .20, respectively) and between the 50% and 75% looking conditions ( $t(33) = 4.471$ ;  $p < .001$ ;  $d = -0.88$  Means: .20, .36; SD: .16, .20, respectively). No significant difference was found between the 25% and the 50% looking conditions ( $t(33) = .279$ ;  $p = .768$ ;  $d = 0.06$ ). A Bonferroni adjustment for multiple comparisons was made for

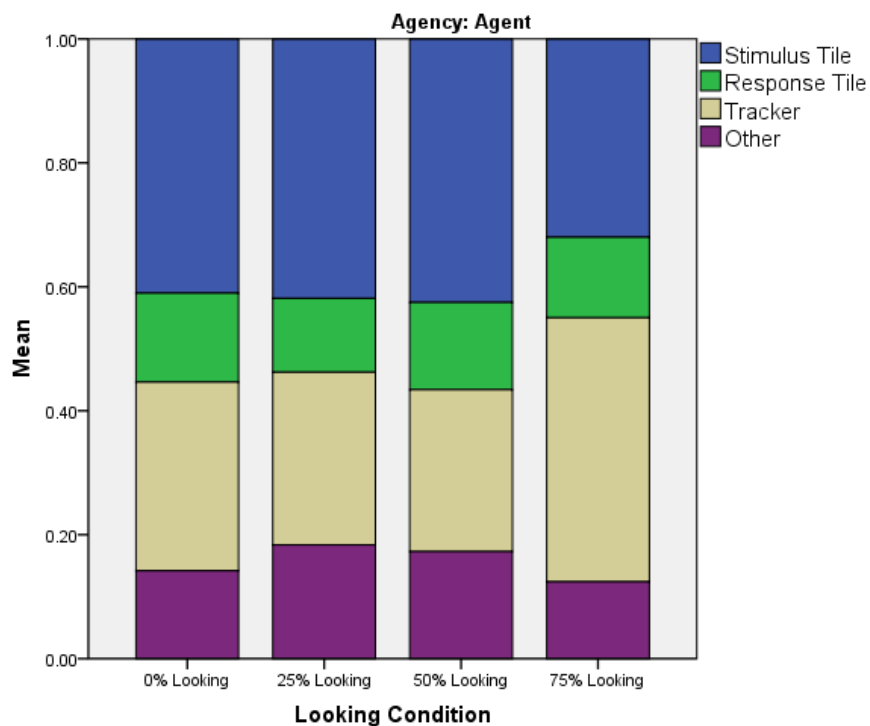


each of the *t*-tests. Although there was a higher proportion of mutual gaze uptake in the Avatar condition than in the Agent condition, this difference failed to reach significance ( $F(1,34) = .990$ ;  $p = .327$ ).

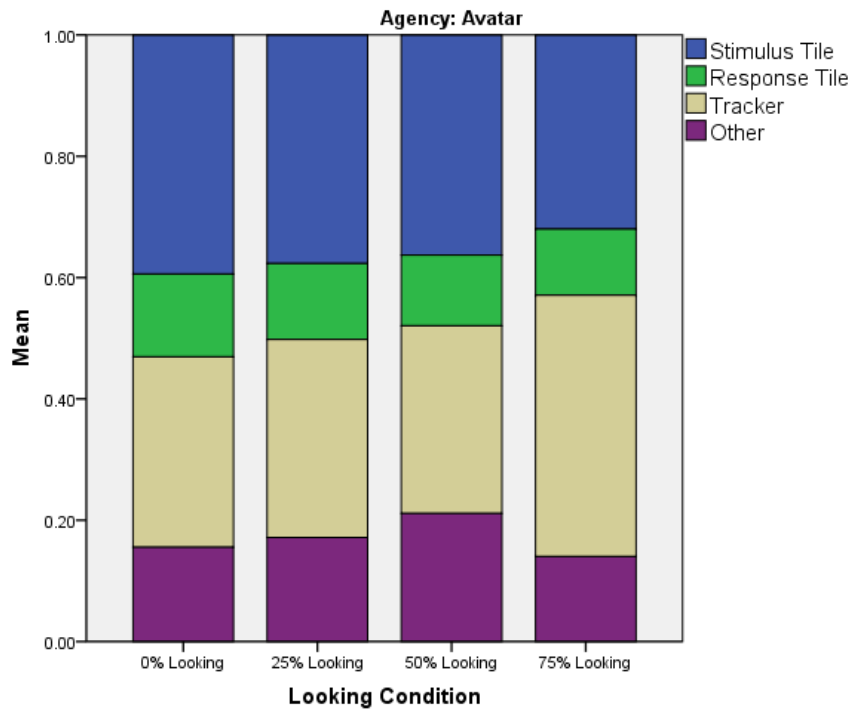
#### 4.3.2.2 Interest Area Analysis

To investigate further the users' looking during the interaction, the screen was divided into interest areas for analysis. The 3 areas of interest were Tracker, stimulus tiles and response tiles, as described in the method section. Any fixations not falling within one of these 3 interest areas were assigned to the category 'other'. The mean proportion of the trial spent looking at each interest area type was compared, as well as the mean number of fixations falling within each interest area.

##### *Dwell Time*



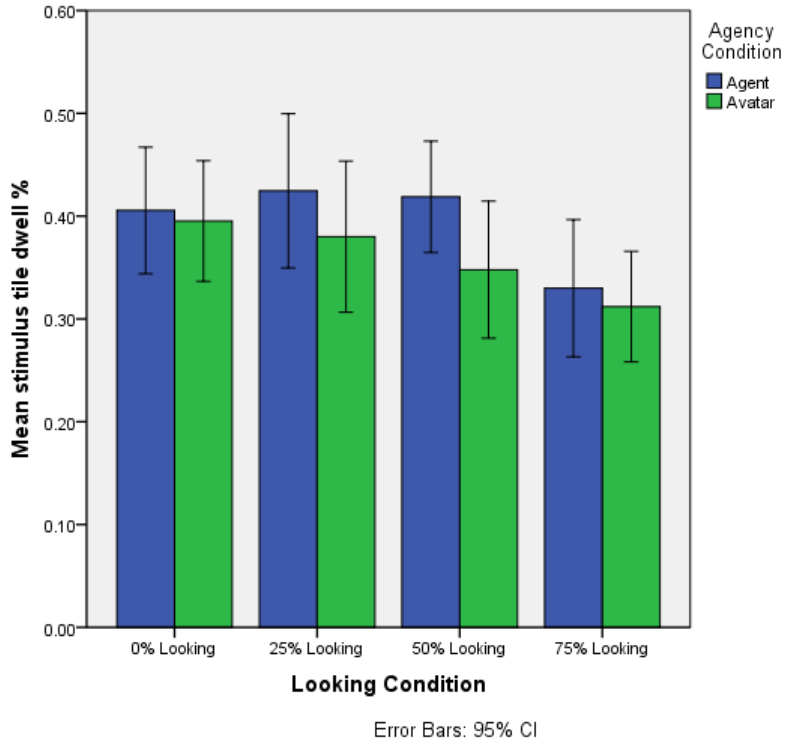
**Figure 17: Mean proportion of trial spent looking at each type of interest area in the Agent condition**



**Figure 18: Mean proportion of trial spent looking at each type of interest area in the Avatar condition**

Figure 17 and Figure 18 show the mean dwell times for each interest area type by agency condition. This was defined as mean proportion of each trial spent looking at stimulus tile, response tile, Tracker and 'other'. The dwell times were compared at each level of the independent variables (looking condition and agency). The figures appear to show a relatively uniform amount of looking at response tiles across all looking conditions and agency. It can be seen that the amount of time spent looking at stimulus tiles and Tracker varies across looking conditions, as well as between the 2 agency conditions. 2x4 mixed ANOVAs (agency x looking condition) were carried out on the proportion of trial spent looking at each IA type to find out if these differences were significant.

## Stimulus Tile

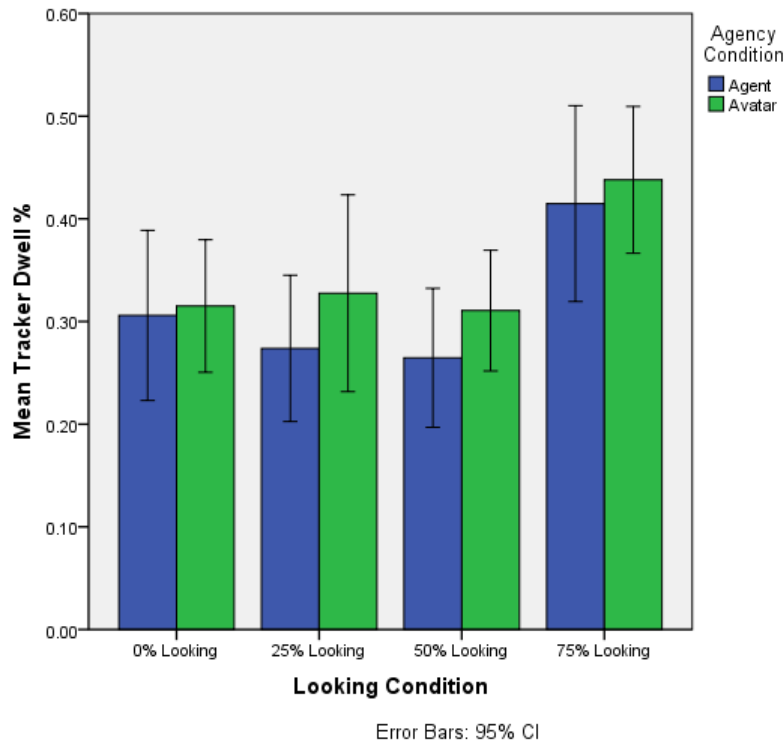


**Figure 19: Mean % of trial spent looking at Stimulus Tiles by Looking Condition and Agency**

No looking condition  $\times$  agency interaction for mean proportion of time spent looking at stimulus tiles was found ( $F(3,108) = 1.019$ ;  $p = .387$ ; see Figure 19). No effect of agency was found ( $F(1,36) = .702$ ;  $p = .407$ ). A main effect of looking condition on mean proportion of time spent looking at stimulus tiles was found ( $F(3,108) = 8.227$ ;  $p < .001$ ;  $\eta_p^2 = .186$ ). A paired samples  $t$ -test revealed that there was significantly less looking at stimulus tiles in the 75% condition than in the 0% condition ( $t(37) = 4.376$ ;  $p < .001$ ;  $d = 0.67$  Means: .32, .40; sd: .12, .12, respectively), significantly less in the 75% than the 25% condition ( $t(37) = 4.199$ ;  $p < .001$ ;  $d = 0.59$  Means: .32, .40; sd: .12, .15, respectively) and significantly less in the 75% than the 50% condition ( $t(37) = 4.120$ ;  $p < .001$ ;  $d = -0.64$  Means: .32, .40; sd: .12, .13, respectively). No significant difference was found between the 0% and 25% conditions ( $t(37) = .248$ ;  $p = .806$ ), the 0% and the 50% conditions ( $t(37) = .390$ ;  $p = .699$ ).

or the 25% and the 50% conditions ( $t(37)=.167$ ;  $p=.868$ ). A Bonferroni adjustment for multiple comparisons was made for each of the t-tests

## Tracker

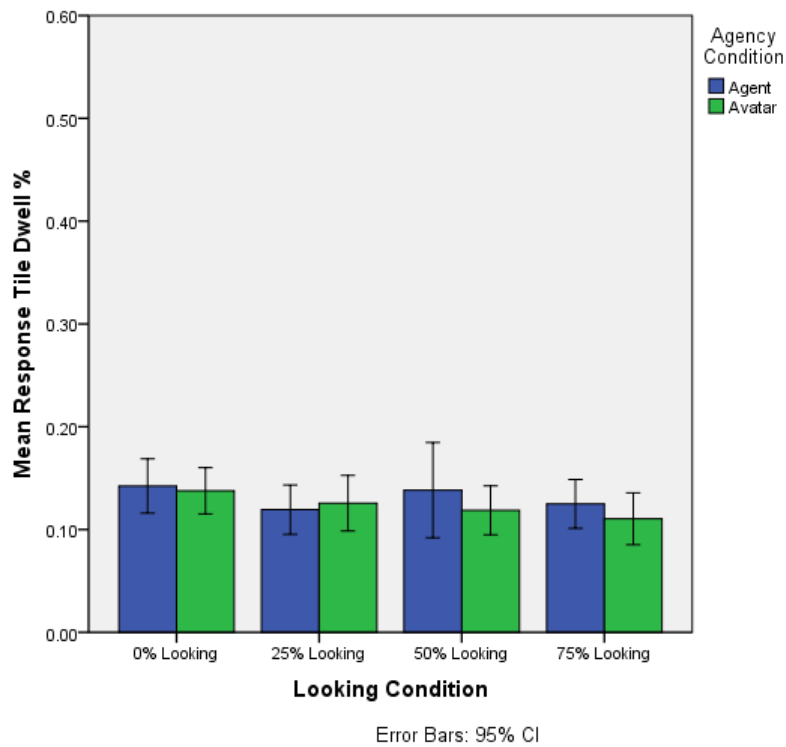


**Figure 20: Mean % of trial spent looking at Tracker by Looking Condition and Agency**

No looking condition x agency interaction was found for the mean proportion of time spent looking at Tracker ( $F(3,108) = .612$ ;  $p=.609$ ; see Figure 20). No main effect of agency was found ( $F(1,36) = .387$ ;  $p=.538$ ). A significant main effect of looking condition on mean proportion of the trial spent looking at Tracker was found ( $F(3,108) = 18.815$ ;  $p < .001$ ;  $\eta_p^2 = .343$ ). A paired samples  $t$ -test revealed that there was significantly more looking at Tracker in the 75% looking condition than in the 0% looking condition ( $t(37)=5.538$ ;  $p<.001$ ;  $d = 0.75$ ; Means: .43, .31; sd: .17, .15, respectively), more looking at Tracker in the 75% condition than in the 25% condition ( $t(37)=7.042$ ;  $p<.001$ ;  $d=0.76$ ; Means: .43, .30; sd: .17, .17, respectively), and more looking at Tracker in the 75% condition than in the 50% condition ( $t(37)=5.928$ ;  $p<.001$ ;  $d= .93$ ; Means: .43, .29; sd: .17, .13, respectively). No

difference was found between the 0% and 25% conditions ( $t(37)=.315$ ;  $p=.754$ ), the 0% and 50% conditions ( $t(37)=1.139$ ;  $p=.262$ ) or the 25% and 50% conditions ( $t(37)=.819$ ;  $p=.418$ ). A Bonferroni adjustment for multiple comparisons was made for each of the  $t$ -tests.

### Response Tiles

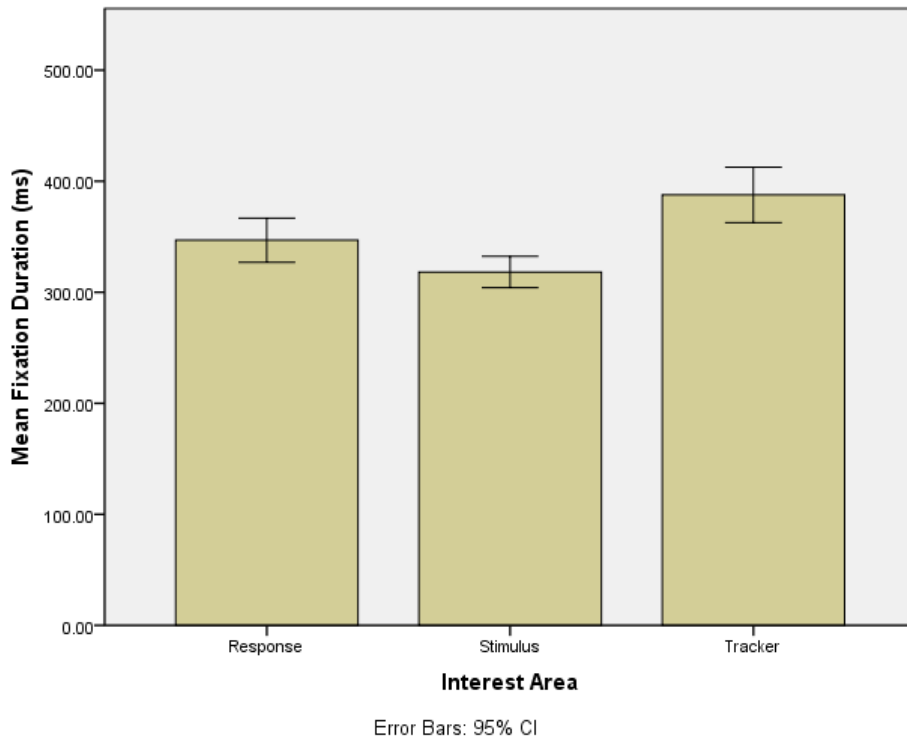


**Figure 21: Mean % of trial spent looking at Response Tiles by Looking Condition and Agency**

Mauchly's Test of Sphericity indicated that the assumption of sphericity for Looking Condition had been violated ( $\chi^2(5) = 13.624$ ,  $p < .05$ ). No looking condition x agency interaction was found for mean proportion of trial spent looking at response tiles ( $F(3,108) = .878$ ;  $p=.436$ , Greenhouse-Geisser adjusted; see Figure 21). No main effect of agency on response tile dwell time was found ( $F(1,36) = .667$ ;  $p=.419$ ), and no main effect of looking condition was found ( $F(3,108) = 1.549$ ;  $p=.215$ , GG adjusted).

### ***Fixation Duration by Interest Area***

Since the previous analysis showed variations in dwell time on Tracker and stimulus tiles, the fixation durations (in ms) within each interest area were analysed to find out if they were driving these differences.



**Figure 22: Mean fixation duration by Interest Area type**

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated,  $\chi^2(2) = 6.053$ ,  $p < .05$ . A main effect of interest area on mean fixation duration was found ( $F(2,64) = 19.629$ ;  $p < .01$ ;  $\eta_p^2 = .380$ , GG adjusted; see Figure 22). The mean fixation duration was found to be significantly longer when looking at Tracker than when looking at response tiles ( $t(16) = 3.061$ ;  $p < .01$ ;  $d = 0.68$  Means: 346.9, 387.7; SD: 58.4, 62.3, respectively). Furthermore, there were significantly longer fixations at Tracker than at stimulus tiles ( $t(16) = 6.47$ ;  $p < .01$ ;  $d = 1.33$ ; Means: 318.2, 387.7; SD: 39.4, 62.3, respectively). No difference was found between the length of fixations when looking at response tiles and when looking at stimulus

tiles ( $t(16) = 2.33$ ;  $p = .033$  (Means: 346.9, 318.2; SD: 58.4, 39.4, respectively). A Bonferroni adjustment for multiple comparisons was made for each of the  $t$ -tests.

### **4.3.3 Task performance and Mutual gaze**

A Shapiro Wilk test for normality discovered that some of the data was non-normally distributed. Both a Pearson's and a Spearman's correlation coefficient were calculated for each comparison and a non-parametric, rather than parametric, correlation coefficient was reported where appropriate. All tests were two-tailed, unless otherwise specified.

#### **4.3.3.1 Overall**

In order to look at the overall effect of mutual gaze on task performance, a correlation analysis was carried out on all of the data collapsed across looking condition and agency.

##### ***Number correct***

The relationship between the number of tasks out of 10 that were correctly completed and the uptake of mutual gaze was examined.

A Pearson's correlation found that the overall relationship between mutual gaze and number of questions answered correctly, when the data was collapsed across IVs, failed to reach significance ( $r = .89$  (34),  $p = .617$ ).

##### ***Reaction Time (RTEnd)***

The time lapsed (in ms) from the end of the instructions (i.e. when the video of IG instructions finished) to when a response tile was selected (RTEnd) was investigated. The relationship between this measure and the proportion of opportunities for mutual gaze taken up (uptake of mutual gaze) was examined. A Pearson's correlation found that the overall negative relationship between RTEnd and proportion of uptake of mutual gaze was found to be non-significant ( $r = -.216$  (17),  $p = .219$ ).

#### **4.3.3.2 Agency**

To find out if there was a difference in relationship between MG and task performance in each agency condition, the data was collapsed across looking condition and examined for the Agent and Avatar conditions separately.

##### ***Number correct***

A Pearson's correlation found that the relationship between mutual gaze and number of questions answered correctly in the Agent condition failed to reach significance ( $r = -.174$  (17),  $p = .503$ ). A Pearson's correlation found that the relationship between mutual gaze and number of questions answered correctly in the Avatar condition also failed to reach significance ( $r = .355$  (17),  $p = .163$ ).

##### ***Reaction Time***

A Pearson's correlation found that the relationship between mutual gaze and reaction time (RTEnd) in the Agent condition was non-significant ( $r = -.074$  (17),  $p = .776$ ). A Pearson's correlation found that the relationship between mutual gaze and reaction time in the Avatar condition also failed to reach significance ( $r = -.304$  (17),  $p(\text{two-tailed}) = .236$ ).

#### **4.3.3.3 Looking Condition**

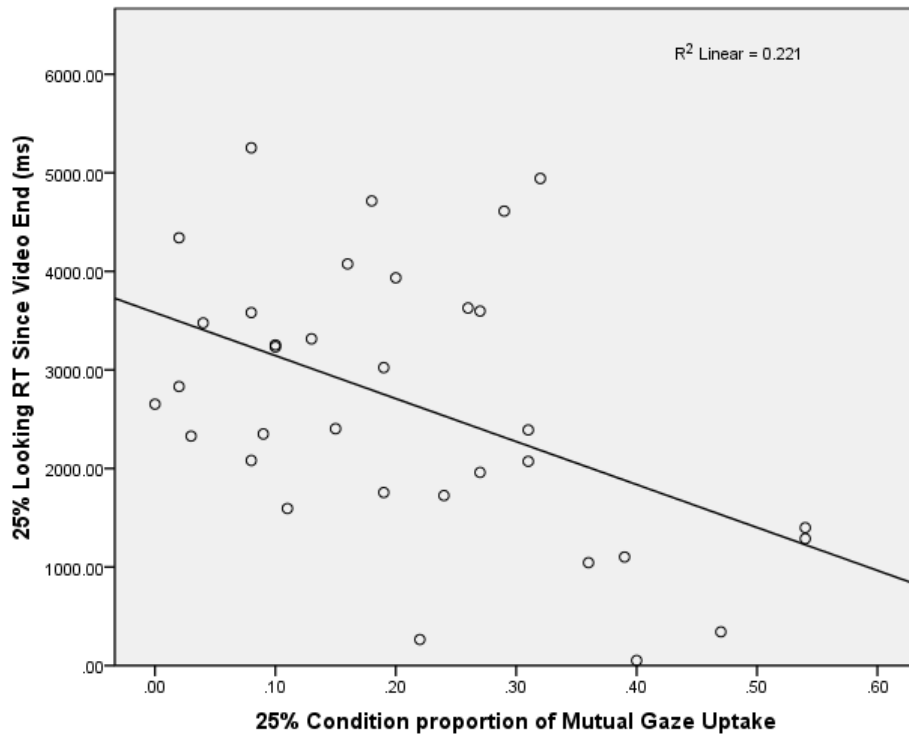
Next, the data was split by looking condition.

##### ***Number correct***

A Pearson's correlation found that the relationship between mutual gaze and number of tasks correctly completed in the 25% looking condition failed to reach significance ( $r = .078$  (34),  $p(\text{one-tailed}) = .331$ ). A Spearman's correlation found that the relationship between mutual gaze and number of tasks correctly completed in the 50% looking condition also failed to reach significance ( $r_s = -.113$  (34),  $p(\text{one-tailed}) = .263$ ), as did the same analysis in the 75% looking condition ( $r_s = -.016$  (34),  $p(\text{one-tailed}) = .465$ ).



## Reaction Time



**Figure 23: Proportion of MG uptake and RT in 25% looking condition**

Since a significant negative correlation between mutual gaze and reaction time had been found in the not-staring condition in experiment 1 (a condition under which Tracker looked at the user during, on average, 27% of the trial), a similar pattern was anticipated in the results from the 25% looking condition, therefore a one-tailed test was used to examine the overall relationship between mutual gaze and reaction time in the 25% looking condition. A Pearson's correlation found a significant negative relationship between the mutual gaze and RTEnd in the 25% looking condition ( $r = -.470$  (34),  $p(\text{one-tailed}) < .01$ ; see Figure 23). Further investigation revealed a significant negative relationship between mutual gaze and reaction time in the avatar condition ( $r = -.491$  (17),  $p(\text{two-tailed}) < .05$ ), but not in the agent condition ( $r = -.344$  (17),  $p(\text{two-tailed}) = .177$ ).

A Spearman's correlation found that the relationship between mutual gaze and reaction time since end in the 50% looking condition failed to reach significance ( $r_s = -.136$  (34),  $p(\text{one-tailed}) = .221$ ).

#### 4.3.4 Questionnaire

In order to investigate users' perceptions of Tracker and the environment, questionnaire was gathered from participants and analysed. The mean responses to each questionnaire item for the Agent and Avatar conditions under each looking condition are shown in Table 2 and Table 3, respectively.

**Table 2: Mean and standard deviation for each questionnaire item by looking condition – Agent.**

Agent Condition	0% Looking		25% Looking		50% Looking		75% Looking	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
How much were you able to control Tracker's behaviour?	1.36	1.22	1.09	1.38	0.95	1.40	1.23	1.11
How responsive was Tracker to the actions you performed?	1.91	1.31	1.73	1.35	1.86	1.49	2.14	1.39
How completely were all your senses engaged in your interaction with Tracker?	1.86	1.17	2.05	1.17	2.36	0.73	2.05	1.33
How much did the appearance and behaviour of Tracker involve you in the interaction?	2.09	1.38	1.82	1.22	1.50	1.44	2.45	0.96
How much did Tracker's verbal instructions involve you in the interaction?	2.86	0.83	2.77	0.92	3.23	0.87	2.95	0.84
How aware were you of events occurring in the real world around you?	1.14	0.89	1.32	0.95	1.27	1.12	1.23	0.97
How aware were you of your display and control devices (mouse and monitor)?	2.00	1.20	1.91	1.11	2.00	1.20	2.27	0.94
How much did your interaction with Tracker seem consistent with your real-world interaction?	1.45	1.06	1.68	1.17	1.64	1.05	1.82	0.96

Were you able to anticipate what would happen next in response to the actions that you performed?	1.23	1.27	1.27	1.24	1.73	1.32	1.45	1.30
How clearly could you understand Tracker's instructions?	3.09	0.92	3.14	0.71	3.41	0.80	3.32	0.65
To what degree did you feel confused or disorientated at the beginning of breaks or at the end of the experimental session?	0.73	0.77	0.73	0.83	0.73	1.16	0.82	0.85
How involved were you in the experience of interacting with Tracker?	2.27	1.28	2.36	1.36	2.55	1.14	2.50	1.26
How distracting was the control mechanism?	0.86	0.94	0.77	0.97	0.91	1.19	0.73	0.94
How quickly did you adjust to the virtual environment experience?	2.64	1.00	2.82	0.66	2.55	0.96	2.91	0.75
How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	2.86	0.99	2.82	0.96	2.59	0.91	2.68	0.95
How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?	1.05	1.09	1.27	1.20	0.82	1.05	1.18	1.14
How much did the control devices interfere with the performance of assigned tasks or with other activities?	0.82	1.05	0.73	0.94	0.82	1.18	0.59	0.80
How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	2.59	1.01	2.73	1.08	2.86	1.08	3.00	0.62

Did you learn new techniques that enabled you to improve your performance?	1.55	1.22	1.45	1.10	1.00	0.76	1.73	1.42
Were you involved in the experimental task to the extent that you lost track of time?	1.41	1.10	1.55	1.06	1.64	1.33	1.91	1.11

**Table 3: Mean and standard deviation for each questionnaire item by looking condition – Avatar.**

Avatar	0% Looking		25% Looking		50% Looking		75% Looking	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
How much were you able to control Tracker's behaviour?	0.86	1.17	0.95	1.40	0.95	1.25	1.05	1.33
How responsive was Tracker to the actions you performed?	2.27	1.28	1.86	1.49	1.86	1.32	2.00	1.54
How completely were all your senses engaged in your interaction with Tracker?	2.59	1.05	2.36	0.73	2.36	1.09	2.50	0.80
How much did the appearance and behaviour of Tracker involve you in the interaction?	1.55	1.41	1.50	1.44	1.64	1.47	1.91	1.63
How much did Tracker's verbal instructions involve you in the interaction?	3.09	0.92	3.23	0.87	3.14	1.04	3.27	0.70
How aware were you of events occurring in the real world around you?	0.95	0.84	1.27	1.12	1.09	0.92	1.14	0.94
How aware were you of your display and control devices	1.82	1.10	2.00	1.20	1.95	1.21	1.82	1.30

(mouse and monitor)?								
How much did your interaction with Tracker seem consistent with your real-world interaction?	1.41	1.10	1.64	1.05	1.36	1.09	1.64	0.90
Were you able to anticipate what would happen next in response to the actions that you performed?	1.68	1.29	1.73	1.32	1.27	1.03	1.77	1.45
How clearly could you understand Tracker's instructions?	3.50	0.67	3.41	0.80	3.32	0.84	3.64	0.49
To what degree did you feel confused or disorientated at the beginning of breaks or at the end of the experimental session?	0.23	0.69	0.73	1.16	0.73	0.98	0.23	0.53
How involved were you in the experience of interacting with Tracker?	2.45	1.26	2.55	1.14	2.55	1.10	2.23	1.38
How distracting was the control mechanism?	0.68	1.04	0.91	1.19	0.73	1.12	0.64	0.85
How quickly did you adjust to the virtual environment experience?	2.64	0.85	2.55	0.96	2.59	0.67	3.09	0.81
How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	2.55	1.06	2.59	0.91	2.50	0.96	2.86	0.89

How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?	1.00	1.11	0.82	1.05	0.91	1.06	0.64	0.85
How much did the control devices interfere with the performance of assigned tasks or with other activities?	0.64	0.85	0.82	1.18	0.77	0.81	0.64	0.73
How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	2.82	0.80	2.86	1.08	2.86	0.83	3.05	0.90
Did you learn new techniques that enabled you to improve your performance?	1.18	0.96	1.00	0.76	1.18	1.18	1.55	1.37
Were you involved in the experimental task to the extent that you lost track of time?	1.41	1.22	1.64	1.33	1.55	1.26	1.50	1.26

A Principle Components Analysis with Varimax rotation was carried out on the 20 Likert scale items making up the adapted social presence questionnaire. This data was gathered from 44 participants, each of whom responded to the set of questions 4 times during the experiment (following interaction with Tracker in each condition). An examination of the Kaiser-Meyer Olkin (KMO) measure of sampling adequacy suggested that the sample was factorable (KMO=.761). Bartlett's test of sphericity was found to be significant ( $\chi^2(190)=1029.162$ ;  $p < .001$ ), indicating that the R-matrix was significantly different from an identity matrix, therefore there was

sufficient correlation between the variables for factor analysis to be suitable. The original questionnaire from which the current questions were taken and adapted showed a 4 factor solution, therefore a 4-factor solution was specified during analysis. The factor loadings after orthogonal rotation of the solution are shown in Table 4. When loadings less than 0.30 were excluded, the analysis yielded a four-factor solution with a simple structure (factor loadings  $\geq .30$ ).

**Table 4: Varimax rotated component loadings for 20 questionnaire items**

	Component			
	1	2	3	4
1 How much were you able to control Tracker's behaviour?	.698			
2 How responsive was Tracker to the actions you performed?	.648	-.260	.147	-.102
3 How completely were all your senses engaged in your interaction with Tracker?	.465		-.106	
4 How much did the appearance and behaviour of Tracker involve you in the interaction?	.409			-.129
8 How much did your interaction with Tracker seem consistent with your real-world interaction?	.527		.435	-.130
9 Were you able to anticipate what would happen next in response to the actions that you performed?	.480		.187	.195
12 How involved were you in the experience of interacting with Tracker?	.729	-.229	.113	
15 How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	.520	-.228	.417	.461
20 Were you involved in the experimental task to the extent that you lost track of time?	.637		-.273	
18 How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	.461	-.439	.117	
13 How distracting was the control mechanism?	-.117	.762		
16 How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?		.574	-.394	.130
17 How much did the control devices interfere with the performance of assigned tasks or with other activities?		.872	-.127	
5 How much did Tracker's verbal instructions involve you in the interaction?	.344		.498	-.335
10 How clearly could you understand Tracker's instructions?		-.135	.709	
11 To what degree did you feel confused or disorientated at the beginning of breaks or at the end of the experimental session?	.148	.247	-.670	
14 How quickly did you adjust to the virtual environment experience?	.420	-.220	.526	.265

6 How aware were you of events occurring in the real world around you?	-.114	.354	.370	
7 How aware were you of your display and control devices (mouse and monitor)?	.324	.198	.108	-.681
19 Did you learn new techniques that enabled you to improve your performance?	.258	.336		.636

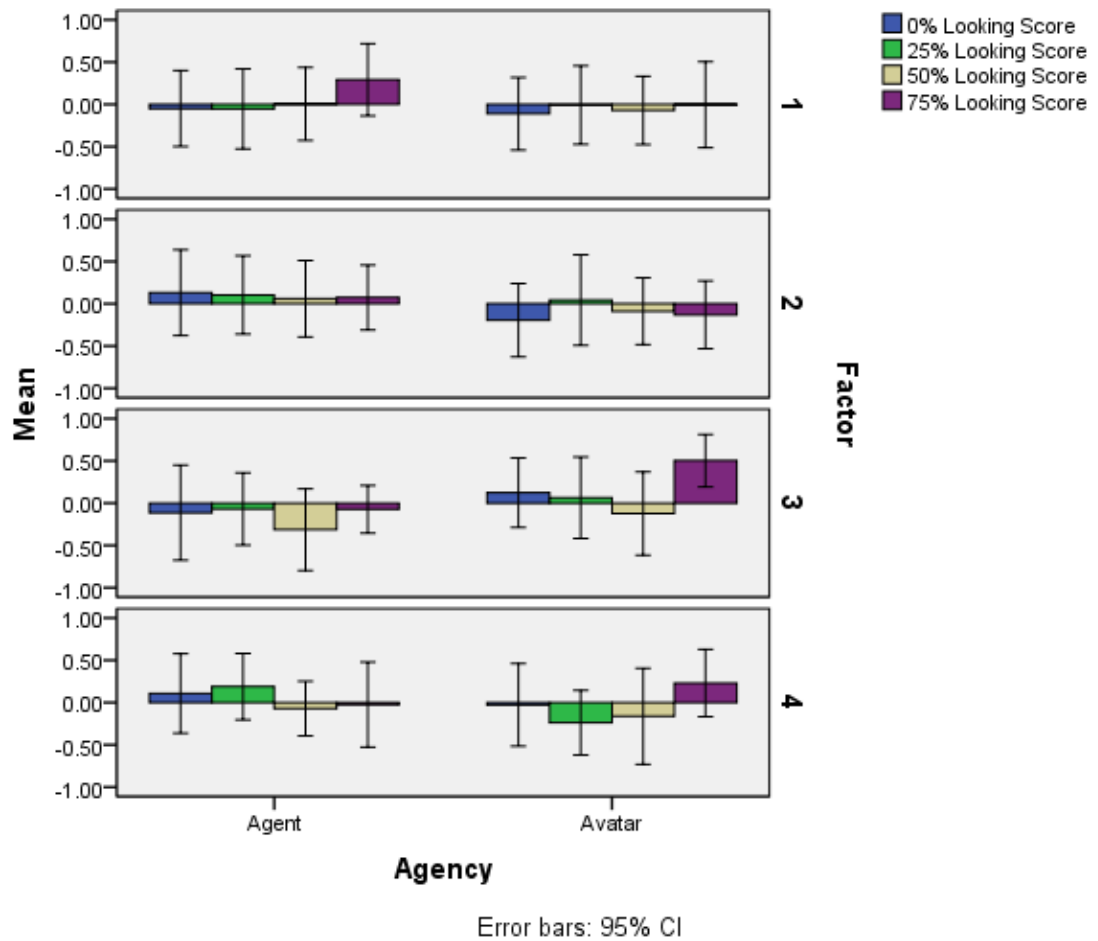
#### 4.3.4.1 The Four Factors

10 items were found to load highly onto factor 1. These can be seen to be related to the user's involvement with Tracker and the environment, specifically in terms of the behaviour of Tracker and how much the user's actions impacted on Tracker and vice versa. This will be referred to as *interface/behaviour*. The second factor appears to be related to the negative aspects of the interface, such as how distracting were the control mechanisms. This will be referred to as *immersion*. The third factor appears to be also related to the user's engagement with Tracker and the environment. These items, however, seem to be less to do with behaviours and how they can be influenced by an interlocutor (as factor 1 appears to deal with) and more to do with thoughts and perceptions of Tracker and the environment, i.e. immersion and presence. This will be referred to as *social interaction/perceptions*. Factor 4 only has 2 items loading highly onto it, so is more difficult to interpret and name, but includes a question about techniques learnt during the interaction, as well as the negative impact of the control devices.

#### 4.3.4.2 Factor Scores

Factor scores were calculated and saved in SPSS during the PCA. These factor scores were used for further analysis.





**Figure 24: Means and Standard Deviations of Factor Scores for each factor by Looking Condition and Agency**

Figure 24 shows the distribution of factor scores for each factor by agency and looking condition, including the interquartile range and 95% confidence intervals. The initial analysis to be carried out was a comparison between the factor scores for each level of the independent variables (looking condition and agency). A 4x2 mixed (within-between) ANOVA was carried out on the factor scores for each factor.

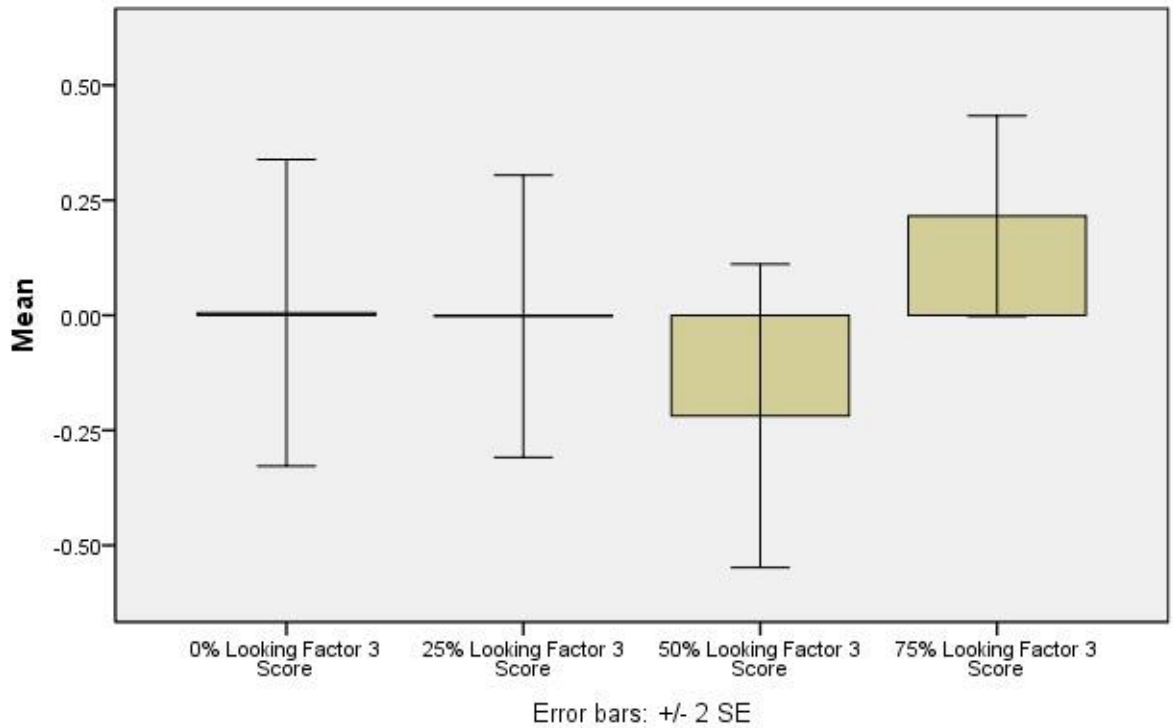
### **Factor 1**

No interaction was found between looking condition and agency was found for the factor 1 factor scores ( $F(3, 126) = 1.014$ ;  $p = .389$ ). No main effect of looking condition ( $F(3,126)=1.914$ ;  $p = .131$ ), or agency (Agency  $F(1,42) = .119$ ;  $p = .732$ ) were found.

## Factor 2

No interaction was found between looking condition and agency for factor 2 ( $F(3, 126) = .485$ ;  $p = .694$ ). No main effect of looking condition ( $F(3, 126) = .396$ ;  $p = .756$ ) or agency ( $F(1, 42) = .455$ ;  $p = .504$ ) were found.

## Factor 3



**Figure 25: Distribution of Factor 3 scores by Looking Condition**

Factor 3 was related to thoughts and perceptions of Tracker and the environment (social interaction/perceptions). Although no looking condition  $\times$  agency interaction was found for factor 3 factor scores ( $F(3, 126) = .845$ ;  $p = .845$ ), a main effect of looking condition was found ( $F(3, 126) = 2.689$ ;  $p < .05$ ;  $\eta_p^2 = .060$ ). A distribution of factor 3 scores by looking condition can be seen in Figure 25. A paired samples t-test for looking condition was carried out on factor 3 factor scores. Bonferroni adjustments were made for multiple comparisons, and all tests were 2-tailed. The results can be seen in Table 4. A significant difference was found between the 50% and the 75% conditions ( $t(43) = 2.89$ ;  $p < .01$  ( $p = .006$ ); means:  $-.22, .22$ ; sd:  $1.09, .72$ ,

respectively;  $d = .48$ ). This indicates that users perceived their interlocutor to be more socially present in the 75% looking condition than in the 50% looking condition. There was no main effect of agency ( $F(1,42) = 1.48$ ;  $p = .231$ ).

**Table 5: Paired samples t-tests for Looking Condition on Factor 3 factor scores**

	0%	25%	50%
25%	$t(43) = .045$ ; $p = .964$		
50%	$t(43) = 1.51$ ; $p = .138$	$t(43) = 1.872$ ; $p = .068$	
75%	$t(43) = 1.19$ ; $p = .239$	$t(43) = 1.463$ ; $p = .151$	$t(43) = 2.89$ ; $p < .01^{**}$ ( $p = .006$ )

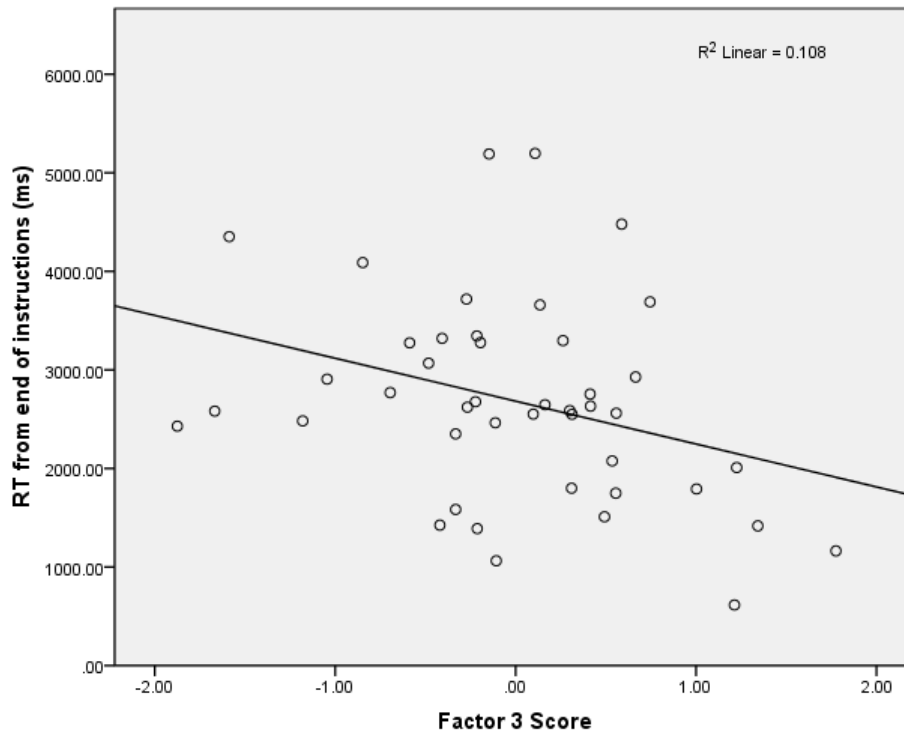
#### **Factor 4**

No looking condition x agency interaction was found for factor 3 factor scores ( $F(3, 126) = 1.871$ ;  $p = .138$ ). No main effect of either looking condition ( $F(3, 126) = .851$ ;  $p = .468$ ) or agency ( $F(1, 42) = .161$ ;  $p = .691$ ) was found.

#### **4.3.4.3 Factor Scores x Reaction Times Correlations**

In order to investigate the relationship between task performance and perceptions of the interaction (social presence perceptions), correlations were carried out on reaction times (from the end from the instructions, as previously examined) and factor scores. Although some of the KS tests of normality were found to be significant, those relating to the variables in which significant correlations were found were non-significant. Coupled with the fact that the accompanying boxplots were found to be relatively symmetrical, it was decided that the parametric test results, i.e. Pearson's correlation coefficients, would be reported.

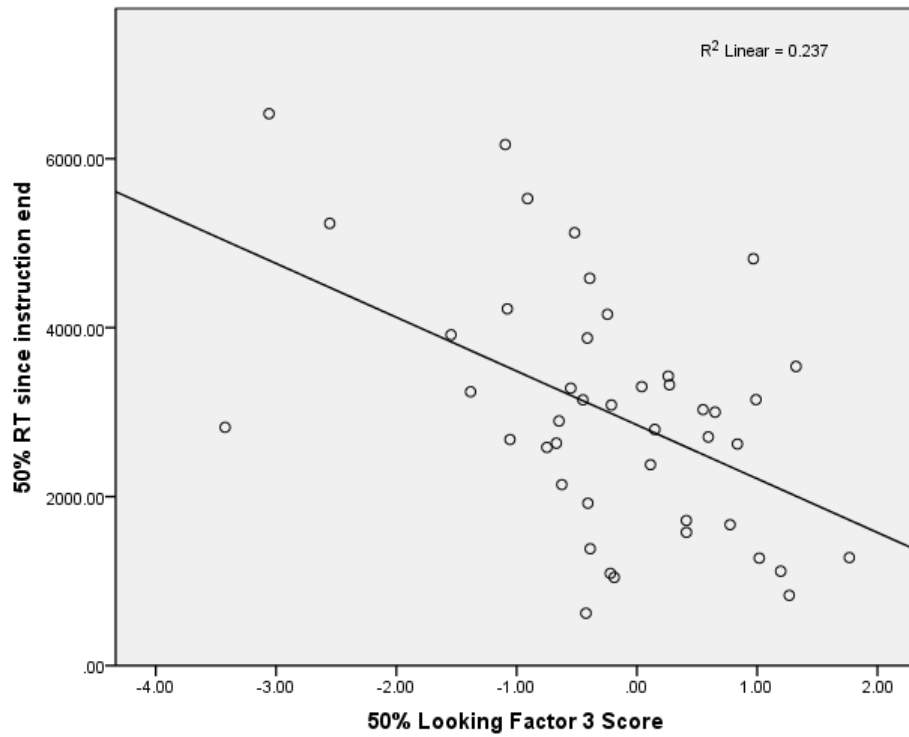
The only factor in which any significant correlations were found was factor 3, the one relating to thoughts and perceptions of Tracker and the rest of the environment, i.e. social interaction/perception (social presence).



**Figure 26: Factor 3 Score and Reaction Time**

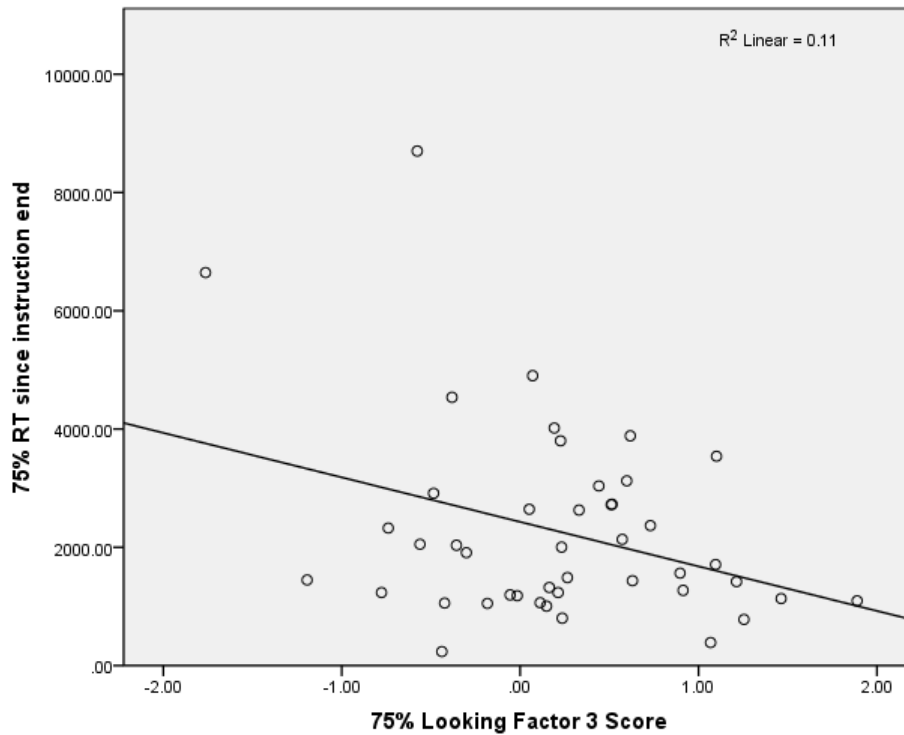
Figure 26 shows the relationship between factor 3 score and RTEnd. A significant negative correlation between factor 3 scores and overall reaction time from end of instructions was found ( $r = -.328$  (44);  $p < .05$ ). When split by agency, the negative relationship in the avatar condition showed a tendency towards significance ( $r = -.407$  (22);  $p = .060$ ), whereas in the agent condition, the negative relationship was found to be non-significant ( $r = -.169$  (22);  $p = .453$ ).

The relationship between factor scores and reaction times were further investigated in each looking condition.



**Figure 27: Factor 3 Score and Reaction Time in 50% Looking Condition**

A significant negative relationship was found between reaction time and factor 3 factor score in the 50% looking condition ( $r = -.487$  (44);  $p < .01$ ; see Figure 27). When split by agency, a significant relationship was found in the agent condition ( $r = -.658$  (44);  $p < .01$ ) but not in the avatar condition ( $r = -.286$  (44);  $p = .197$ ).



**Figure 28: Factor 3 Score and Reaction Time in 75% Looking Condition**

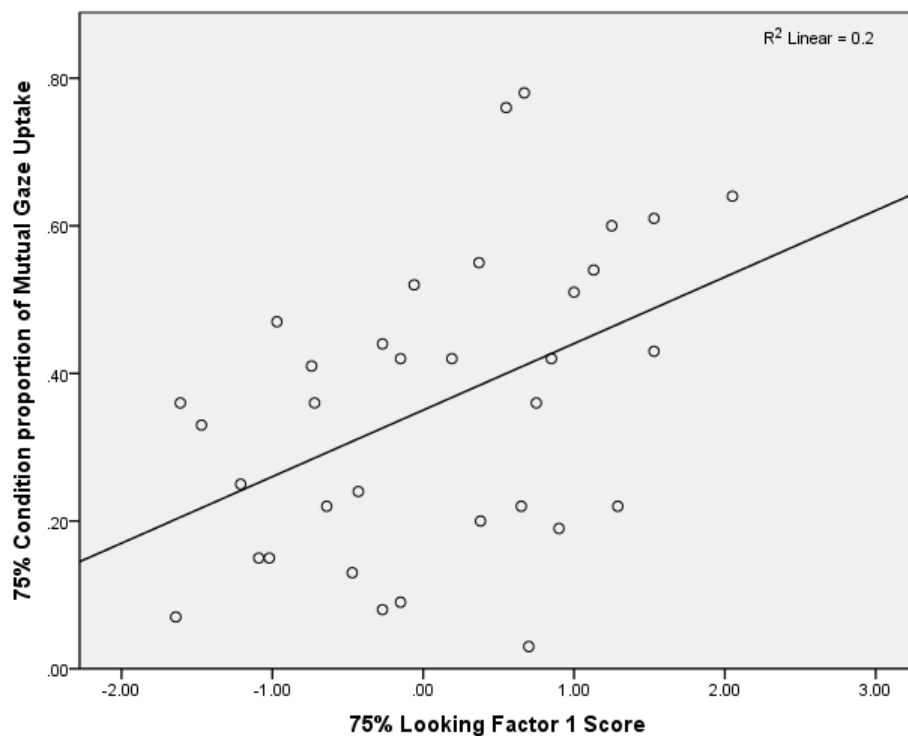
A significant negative relationship was found between factor 3 factor score and reaction time in the 75% looking condition ( $r = -.332$  (44);  $p < .05$ ; see Figure 28). When split by agency, however, the relationships were found to be non-significant (agent:  $r = -.247$  (44);  $p = .268$ ; avatar:  $r = -.215$  (44);  $p = .337$ )

#### **4.3.4.4 Factor Score x Mutual Gaze Correlations**

To establish if there was a relationship between mutual gaze and perceptions of Tracker and the environment (social presence and presence), correlations were investigated between the factor scores for each looking condition and the proportion of mutual gaze that was taken up by the user (proportion of mutual gaze uptake). Where the K-S tests of normality were found to be significant, non-parametric Spearman's correlation coefficients are reported, rather than Pearson's correlations, which are reported for the data found to be normally distributed. All tests are two-tailed.

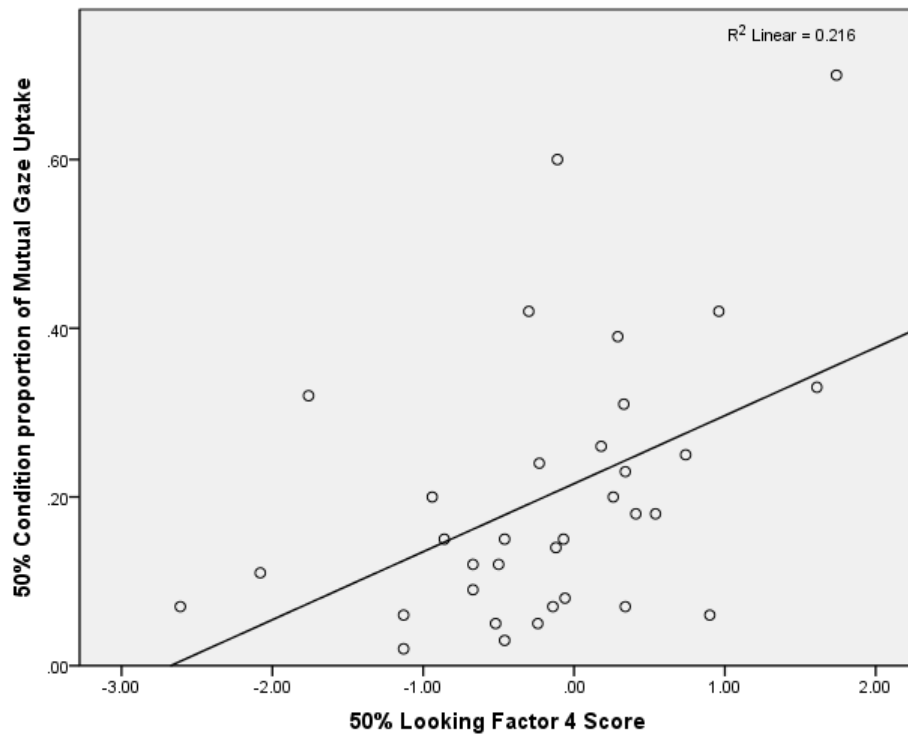
There were no significant correlations between the overall factor scores and the overall mutual gaze uptake for any of the 4 factors. To investigate further, the data was split into looking conditions.

When split into looking conditions, the two factors within which significant correlations with mutual gaze uptake were found were factor 1 (*interface/behaviour*) and factor 4. The questions which load highly onto factor 1 are related to the user's involvement with Tracker and the environment, specifically in terms of the behaviour of Tracker and how much the user's actions impacted on Tracker and vice versa (e.g. How much were you able to control Tracker's behaviour? How responsive was Tracker to the actions you performed?). Factor 1 is fairly similar to factor 3 (*social interaction/perceptions*) but focuses more on behaviours and the environment, rather than perceptions and Tracker. The questions which load highly onto factor 4 are related to immersion and comprehension (e.g. To what degree did you feel confused or disorientated at the beginning of breaks or at the end of the experimental session? How clearly could you understand Tracker's instructions?).



**Figure 29: Factor 1 score and Mutual Gaze Uptake in 75% Looking Condition**

Figure 29 shows the relationship between factor 1 factor score and proportion of mutual gaze uptake in the 75% looking condition. A significant positive correlation was found between the two variables ( $r = .447$  (34);  $p < .01$ ). When split by agency, the relationship was found to show a tendency towards significance in the Agent condition ( $r = .471$  (17);  $p = .056$ ), and was non-significant in the Avatar condition ( $r = .459$  (17);  $p = .064$ ).



**Figure 30: Factor 4 Score and Mutual Gaze Uptake in 50% Looking Condition**

Figure 30 shows the relationship between factor 4 score and uptake of mutual gaze within the 50% looking condition. A significant positive correlation was found between these variables ( $r_s = .461$  (34);  $p < .01$ ). When split by agency, a significant positive correlation was found in the Avatar condition ( $r_s = .560$  (17);  $p < .05$ ), but the relationship in the Agent condition failed to reach significance ( $r_s = .428$  (17);  $p = .087$ ).

Since a significant negative relationship was found between reaction time and factor 3 score in the 75% looking condition, it was anticipated that there would be some



sort of relationship between mutual gaze uptake and factor 3 factor score in that looking condition. The positive relationship between mutual gaze uptake and factor 3 score in the 75% looking condition was found to be non-significant ( $r = .101$  (34);  $p = .570$ ). When split by agency, the correlations again did not reach significance (Agent:  $r = .219$  (17);  $p = .397$ ; Avatar:  $r = -.013$  (17);  $p = .959$ ).

#### **4.3.4.5 Mutual gaze, reaction time and factor score as predictors of Agency**

In order to investigate more fully the relationship between mutual gaze, reaction time, factor score and agency, a binary logistic multiple regression was carried out. The predictors for this analysis were reaction time (ms), the proportion of mutual gaze uptake (as previously defined), factor score for each factor (obtained from the principle components analysis). The binary outcome variable was Agency. The null outcome group (0) was agent, and the alternative outcome (1) was avatar. The aim of this analysis was to establish to what extent each of the predictors affected the outcome, and therefore investigate the magnitude of the relationship between each predictor and the outcome. The data was initially analysed collapsed across looking conditions. Following this, the data was split by looking condition (data from the 0% looking condition was not included, since there was no mutual gaze in that condition) to find out if there were differences in relationship between the predictors and outcome between the groups, as previously discovered in the correlation analyses.

A binary logistic regression was carried out on these 4 groups (data from all looking conditions, from 25% looking, 50% looking and from 75% looking), using the forced entry method.

#### **All looking conditions**

A test of the full model against the model containing only a constant was found to be statistically significant, indicating that the addition of one or more of the predictors would significantly increase the model's ability to predict whether the participant had been assigned to the agent or avatar condition ( $\chi^2$  (6) = 15.8;  $p < .05$ ). A Nagelkerke  $R^2$  value of .192 indicates a fairly weak relationship between predictors and agency grouping. Prediction success was 65.7% (64.7% in agent and 66.7% in avatar). The

Wald criterion indicates that only reaction time made a significant contribution to the prediction of agency ( $\chi^2(1)=10.64$ ;  $p<.01$ ). Mutual gaze, Factor 1, Factor 2, Factor 3 and Factor 4 scores' contributions were found to be non-significant. Exp(B) value indicates that when reaction time is increased by 1ms, the odds ratio is .99 times as large, indicating that likelihood of the participant being in the agent condition decreases by .11 (CI: .999, 1.0).

### **25% looking**

When looking at the data in the 25% looking condition only, a test of the full model against the model containing only the constant was not found to be statistically significant, indicating that the addition of one or more of the predictors would not significantly increase the model's ability to predict whether the participant had been assigned to the agent or avatar condition ( $\chi^2(6) = 8.20$ ;  $p = .224$ ). The only predictor whose contribution was anywhere near significant was factor 4 factor score ( $\chi^2(1)=3.03$ ;  $p=.08$ ). None of the other predictor variables were found to make a significant contribution to the model.

### **50% looking**

In the 50% looking condition, a test of the full model against the null model was found to be non-significant ( $\chi^2(1)=9.14$ ;  $p=.166$ ). The Wald criterion, however, indicates that reaction time did make a significant contribution to the prediction of agency ( $\chi^2(1)=4.44$ ;  $p < .05$ ). Mutual gaze, Factor 1, Factor 2, Factor 3 and Factor 4 scores' contributions were found to be non-significant. Exp(B) value indicates that when reaction time is increased by 1ms, the odds ratio is .99 times as large, indicating that likelihood of the participant being in the agent condition decreases by .11 (CI: .99, 1.0).

### **75% looking**

In the 75% looking condition, a test of the full model against the null model was found to be significant ( $\chi^2(6) = 14.19$ ;  $p < .05$ ). A Nagelkerke  $R^2$  value of .455 indicates a far stronger relationship between predictors and agency grouping (as compared with the data collapsed across looking conditions, where  $R^2=.192$ ). Prediction success was 73.5% (70.6% in agent and 76.5% in avatar). The Wald

criterion indicates that reaction time made a significant contribution to the prediction of agency ( $\chi^2(1)=5.248$ ;  $p<.05$ ). Factor 3 score was found to make a marginal contribution to the prediction ( $\chi^2(1)=3.729$ ;  $p=.053$ ). Mutual gaze, Factor 1, Factor 2, and Factor 4 scores' contributions were found to be non-significant. Exp(B) value indicates that when reaction time is increased by 1ms, the odds ratio is .99 times as large, indicating that likelihood of the participant being in the agent condition decreases by .11 (CI: .99, 1.0). The Exp(B) value indicates that when factor 3 score increases by 1, the odds ratio is 5.450 times as large. The confidence interval, however, crosses 0 so no statistical inferences can be drawn (CI: .98, 30.47).

Overall, it can be seen that this model confirms the previous RT x agency effect, and shows that reaction time is better predictor of agency than any of the others. Most interestingly, RT is able to predict agency better than questionnaire responses, which is counter-intuitive since the questionnaire was designed specifically to detect variations in social presence.

The next chapter will return to the experimental design for both studies, outlining the decision-making process, as well as discussing challenges encountered, and how they were overcome.

## 5 Chapter 5: Experimental Design

At the start of the period of study that culminated in this thesis, there was no precedent for investigating gaze, social presence and task performance within a virtual environment, therefore there was no existing model to follow. As such, the tasks, environment and experimental design were entirely novel. This meant that each separate aspect of each experiment, such as the agent's behaviour, the experimental room and the tasks themselves, had to be entirely designed, created and developed, all of which was extremely time consuming. The methodological aim, as presented earlier in the thesis, was to design a platform suitable for answering the research questions. This chapter addresses this methodological aim, and focuses on the inspiration behind, and reasons for, each design decision in both the human-human and human-agent studies. It also outlines each of the steps involved in creating the experiments and covers some of the challenges encountered during the process.

### 5.1 *Exploratory Study*

This section is based, in part, on a paper presented at ETRA 2008 (Dalziel-Job et al., 2008).

The paradigm used to examine the relationship between mutual gaze, task performance and social presence was based on an exploratory study into how individuals use information from ECAs. It is worth describing it here, since it provided a basis for the design of the paradigms for the current studies (Dalziel-Job et al., 2008). The exploratory study served two purposes: a methodological component and an experimental component. The methodological component aimed to develop a paradigm for study of social interaction within virtual environments. More specifically, a paradigm that would allow the recording and analysis of users' eye movements while they carried out tasks within a virtual environment, assisted by a computer-controlled avatar. One of the methodological issues encountered at the time of the exploratory study, which affected empirical studies of people interacting with ECAs, was that much previous work had focused on systems that were not easily adaptable to domains other than those for which they were developed (Rickel

& Johnson, 1998). Thus, there was a need to develop a paradigm for investigating interaction with ECAs which supported a suitable level of adaptability.

The experimental aim of the exploratory study was to examine how people use information presented by avatars when carrying out tasks in Second Life (SL). The main hypothesis was inspired by a previous investigation in which it was found that a user looks at Genie, an ECA, only 0.1% of the time when carrying out a route-following task, even though that agent is providing visual cues that could help the user to complete the task (Dalziel-Job et al., 2006). The question was, is it possible to make the user pay more attention to an agent that was assisting them in a task? To do this, the agent presented the user with a series of tasks, including sequences to complete, as well as sets of shapes from which to select a mirror image or rotation of a target shape. During some of the trials, the agent both named and pointed at the target shape (redundant pointing), and in others the agent simply pointed and referred to 'this one' (non-redundant pointing). There was a third condition in which the user only heard the voice of Tracker, the computer-controlled avatar (agent) robot giving instructions. An example of the redundant and non-redundant trials can be seen in Figure 31 and Figure 32, respectively. The aim was to establish if the non-redundant pointing would compel the user to look more at the agent than the redundant pointing would. The study also aimed to explore the relationship between looking at an agent, task performance and self-report measures.

An experimental tower was built within Second Life. A tower was chosen, as opposed to a building with three adjacent rooms, because of the limited amount of real estate space allocated for the experiment. The tower consisted of 3 separate rooms built on top of each other. Each participant was presented task-completion instructions by Tracker in three different forms: voice only, visible but providing redundant visual information (see Figure 31), visible and providing non-redundant visual information (see Figure 32). This produced 3 different Tracker conditions: *invisible*, *redundant* and *non-redundant*.

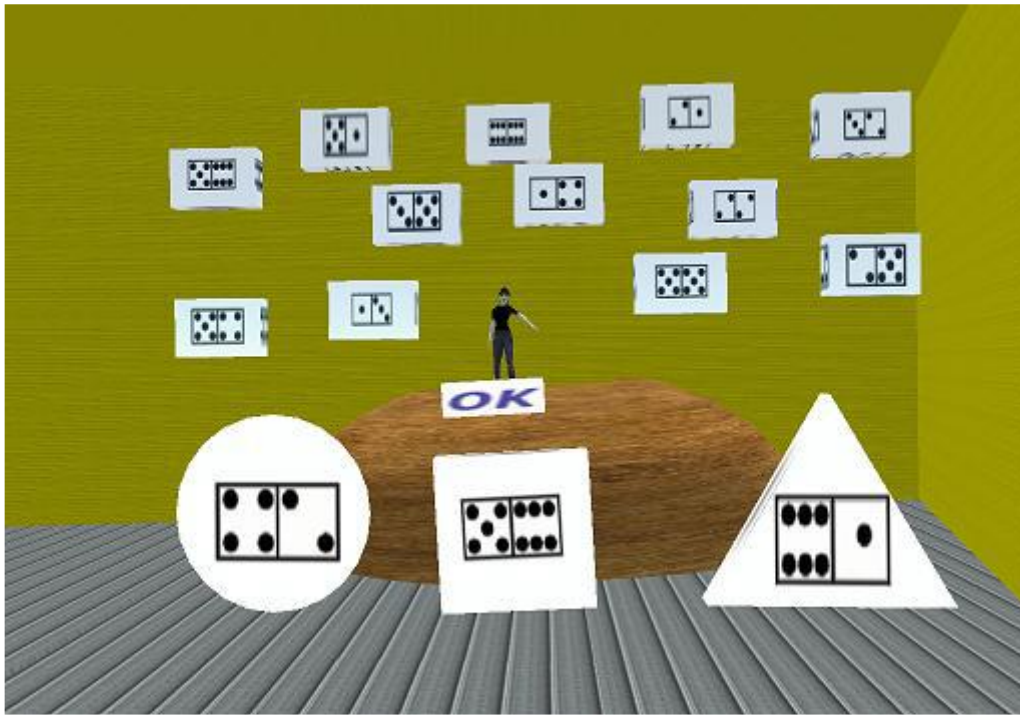


Figure 31: Redundant condition: “From the wall behind me, select a domino with the same total number of dots as the one on the triangle”.

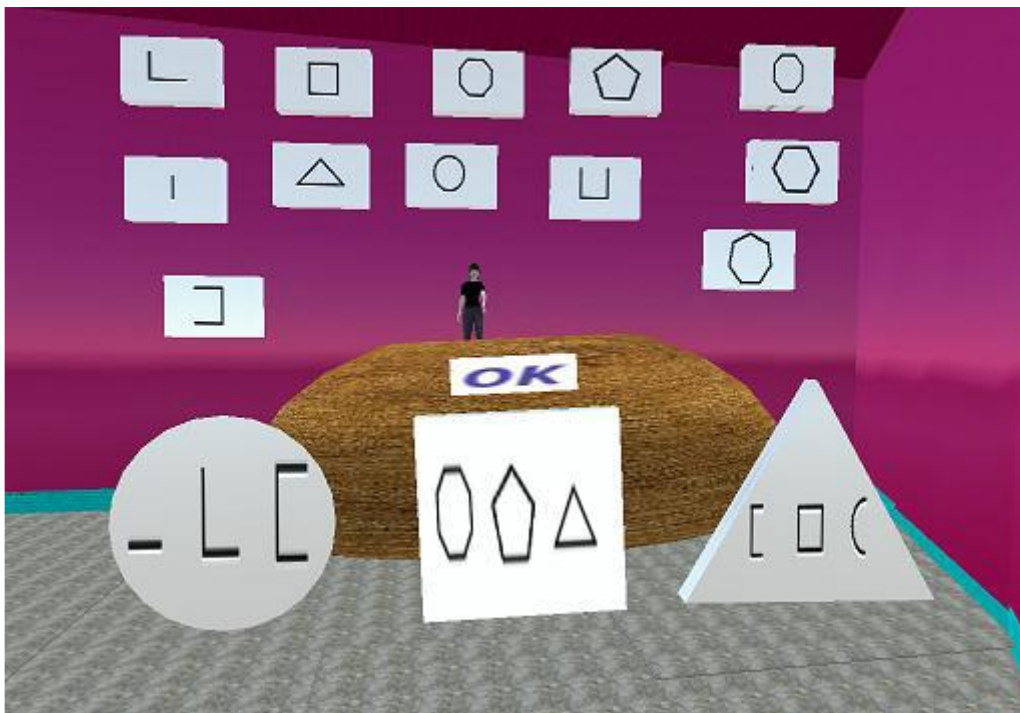


Figure 32: Non-redundant condition: “From the wall behind me, select a shape that you think best completes *this* sequence”.

Owing to the exploratory nature of the eye tracking study within SL, it emerged that there was a lack of dedicated software suitable for carrying out coding of the data. This was an issue which recurred in the following studies. The data was annotated using NITE XML toolkit (Carletta, Evert, Heid, & Kilgour, 2005). The dependent variables were: time spent looking at Tracker while he was giving instructions; number of tasks correctly completed; time taken to complete the tasks; and subjective ratings of the perceived helpfulness of Tracker for each task.

Supporting the results of the Dalziel-Job et al. (2006) Genie study, it emerged that participants looked very little at Tracker while he was talking, in both of the visible conditions. On average, less than 5% of this time was spent looking. There was significantly more looking at Tracker in the non-redundant than redundant condition. There was also a significant negative correlation between the amount of looking at Tracker and the number of correct targets selected in the redundant condition; more redundant looking at the agent was associated with poorer task performance. This pattern was not found in the non-redundant condition. Participants reported Tracker to be more helpful in the non-redundant condition than in either the redundant or invisible conditions.

This study confirms that very little time is spent looking at an agent that is giving instructions, but that this small amount of looking can be increased by ensuring the agent holds task-critical visual information. It was found that, in the redundant condition, looking at an agent appears to be detrimental to task performance. This suggests that the question of whether an ECA should be included in an interface may depend on the type of task to be performed.

Although this exploratory study was outwith the period of research that relates to this thesis, it is important to cover it here, as it formed the basis of the experiments that followed, and some of the later design decisions were informed by those in the exploratory study. The exploratory study showed Second Life to be an extremely versatile and useful platform for designing and running eye tracking studies. Furthermore, it provided useful foundations for developing an experimental paradigm in SL that would allow researchers to evaluate systems that include, or consist of, ECAs assisting users in completing a task.

Inspired by the Dalziel-Job et al. (2008) pointing study, a paradigm was further developed for the current studies, involving an instruction giver assisting a user (whose eye movements were being recorded) in a task-oriented interaction within Second Life. While the studies covered in this thesis were based on the paradigm described above, it must be noted that Second Life proved to be an environment which was updated frequently, meaning that none of the code controlling the above experiment could be re-used for either of the subsequent studies. Indeed, some updates required fairly extensive editing of the code that controlled Tracker's behaviour and the running of the experiments, sometimes during the design process (or even the experimental run). Additionally, updating technology outside Second Life allowed for different data collection and annotation methods for the subsequent two experiments.

## ***5.2 Design Requirements of the Human-Human Experiment***

The first experiment involved two participants – an Instruction Giver (IG) and Instruction Follower (IF) – collaborating via avatars to complete arithmetic tasks. Following on from the exploratory study, a paradigm for measuring the eye movements of two individuals during a task-based interaction was required. The participants were to communicate through avatars in Second Life to investigate mediated social interaction, and the effects of different avatar/agent behaviours on perceptions of and behaviours towards that virtual human. Within the paradigm, users needed to have a feeling of being collocated with a conversational partner, and a sense of 'being together', therefore a realistic 3D virtual environment was required. Participants needed to be able to manipulate objects within the environment, and be able to observe changes in response to those manipulations. Both users had to be able to see the same experimental stimuli, as well as each other, and how the other was interacting with those stimuli (for example, looking at a specific object within the environment). A paradigm within Second Life was able to fulfil all of these criteria. From a practical point of view, as our previous work has shown, Second Life is freely available (or inexpensive) and a relatively straightforward environment in which to set up an experimental room.

The ability to record time-stamped button presses and behaviours was also important.



It would be prudent to cover here some of the steps required to extract the output from the experiments and transform it for analysis. Second Life produces a log file for each interaction which does just that – it records each input into the environment, and behaviours / texture changes in response to that behaviour. This meant that time-stamped responses, in the form of eye movements and mouse clicks could be recorded and exported into software for annotation and analysis. The mouse clicks were saved by Second Life in the form of log files, and the instruction follower eye movements were retrieved from the eyetracking PC in the form of Eyelink Data Files (EDFs). Each EDF was converted into an ascii file using a perl script. All other occurrences within each trial – texture changes, Tracker behaviour, onset of audio instruction files etc. – were also recorded and exported from Second Life in the form of log files. Finally, a video of each trial was captured and saved using Camtasia screen recorder (see <https://www.techsmith.com/camtasia.html>). Many of these videos required a significant amount of editing before they could be interpreted in any meaningful way (they were used in the human-human experiment analysis to extract task performance data and assign regions of interest for eye movement analysis). All of the resulting data files for every trial were amalgamated into one spreadsheet, the conditions were added, and the spreadsheets were further edited where necessary (the eye movement files, for example, had a different time format to the other data files, so all eye movement data had to be transformed). Finally, the spreadsheets for all of the trials were combined to produce one large dataset for analysis of all participants.

The experimental room and stimuli needed to be built to specific requirements. The experiments needed to be able to be initiated by typing in simple commands, and the set of trials in each block had to run consecutively, without input from the experimenter. Each user's avatar had to be sitting, thus eliminating the possibility that they would move around the environment. This ensured that each participant viewed the instructions and stimuli in exactly the same way.

As previously mentioned, the constant updating within the Second Life environment meant that the code that controlled Tracker during the exploratory study no longer worked by the time the human-human experiment was run, necessitating a new

controller to be written. One of the issues with the exploratory study was that Tracker appeared fairly small on the screen. The human-human study required the instruction follower to be able to see the eye movements of the other's avatar (see Human-Human Method section for explanation of how the eye movements of the instruction giver's avatar were controlled). This meant that the avatar on each user's screen should appear large enough to be able to discriminate between its looking at different objects. It was therefore decided that the avatars should sit close enough together to maximise the probability of the IF seeing the non-verbal behaviours of the other's avatar; the distance between the avatars was adjusted until the optimum position for IF viewing of IG behaviour was reached (as judged by the experimenters).

### ***5.3 Additional requirements for the Human-Agent Experiment***

In the second experiment, one user interacted with Tracker, a programmed avatar (i.e. an agent) to complete arithmetic tasks. The experimental design requirements were similar to those of the human-human study, with some additions. The agent needed to:

- Have realistic gaze behaviour, manipulated by the experimenter and scripted to change in response to actions by the user.
- Present task instructions verbally in the form of a pre-recorded audio file. These instructions should be triggered in response to actions carried out by the user.
- Behave, and more specifically, speak in a believably human-like fashion; Tracker should be able to lip sync (i.e. the mouth movements should correspond in a behaviourally realistic fashion with the audio instructions). This proved to be one of the most challenging aspects of the experimental set-up.

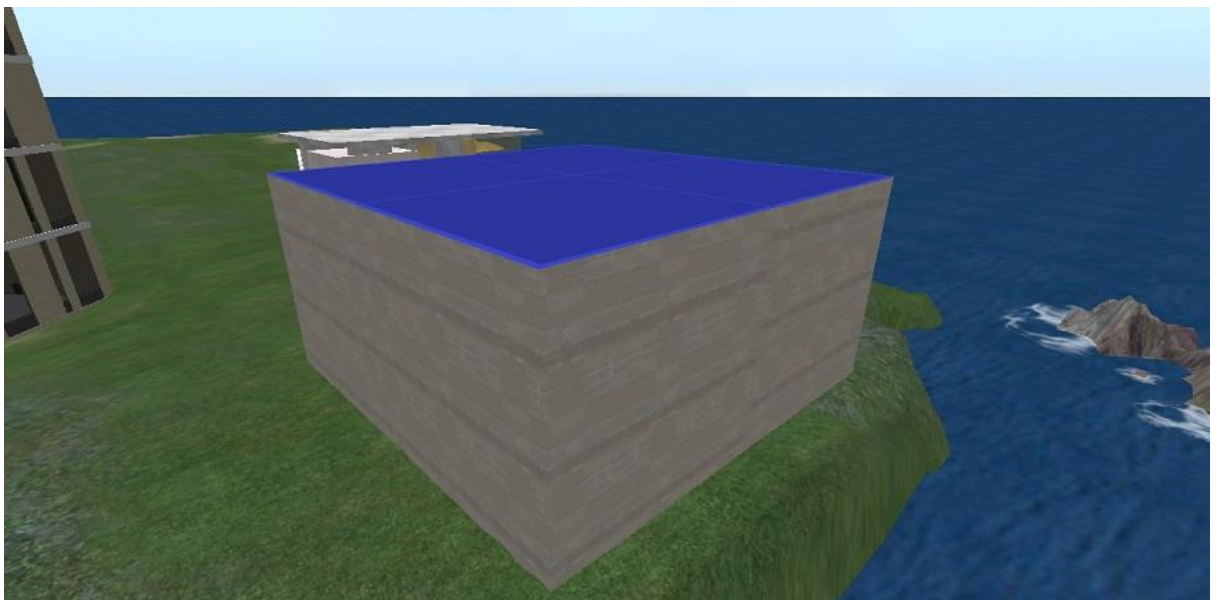
A set of questionnaire items were to be presented after each block of trials. These needed to be answered in-world to try and maintain as much of a feeling of presence as possible, and the agent was not to be present during the presentation of questions.

## **5.4 Human-Human experiment: Independent variables**

Instruction giver looking behaviour was manipulated to produce two contrasting looking behaviours – staring and not staring. One of these was intended to mimic naturalistic looking behaviour; the instruction giver was free to control how much their avatar looks at instruction follower. This was also intended to produce some idea of what is the typical amount of looking by IG and IF during instruction giving, for this particular task, to be incorporated into the Human-Agent study. The staring behaviour was designed to establish whether staring by one person at another is, indeed, an appropriate way to investigate mutual gaze or if it will, in fact, discourage mutual gaze between a dyad.

A within participants design was decided upon to minimise variation due to individual differences, and the tasks were counterbalanced. The staring condition was presented first, as it was deemed to be more straightforward to instruct the IG to perform additional behaviours in the second block – i.e. manipulate the avatar’s gaze – than to remove behaviours they may already have become used to performing – i.e. maintain the gaze towards the other avatar (see the Procedure in the method section for a description of how this manipulation was achieved).

## **5.5 Building an experiment within Second Life**



**Figure 33: Experimental room**

In order to build an experimental room within Second Life, an area within the Virtual University of Edinburgh (VUE) was allocated as experimental space. To create the building, the floor and ceiling and each wall was created out of a basic shape (or *primitive*). Each prim's 3-dimensional size, rotation and position was then edited, as well as its texture (the pattern on it). The prims were then placed together to form an experimental room (see Figure 33). The room was then made 'closed' to avoid the appearance of unwanted visitors during experimental runs. This meant that only the avatars of individuals who had been invited could enter the room. The exceptions to this were on two occasions: the first was where the Virtual University of Edinburgh was filled with bouncing bubbles, which managed to penetrate the experimental room, and the other was when several unknown avatars suddenly appeared in the room. The reasons for these two events remain unclear, and the data collected during the experiments which were running at the time was discarded.

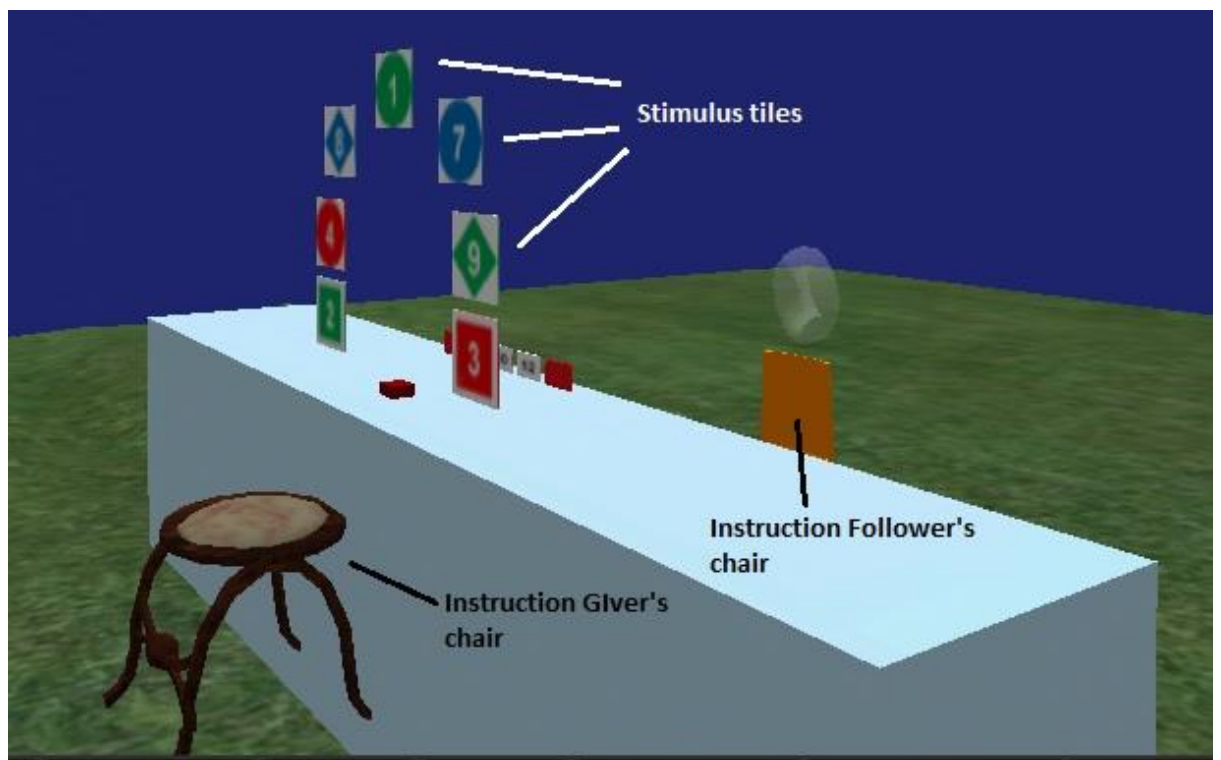
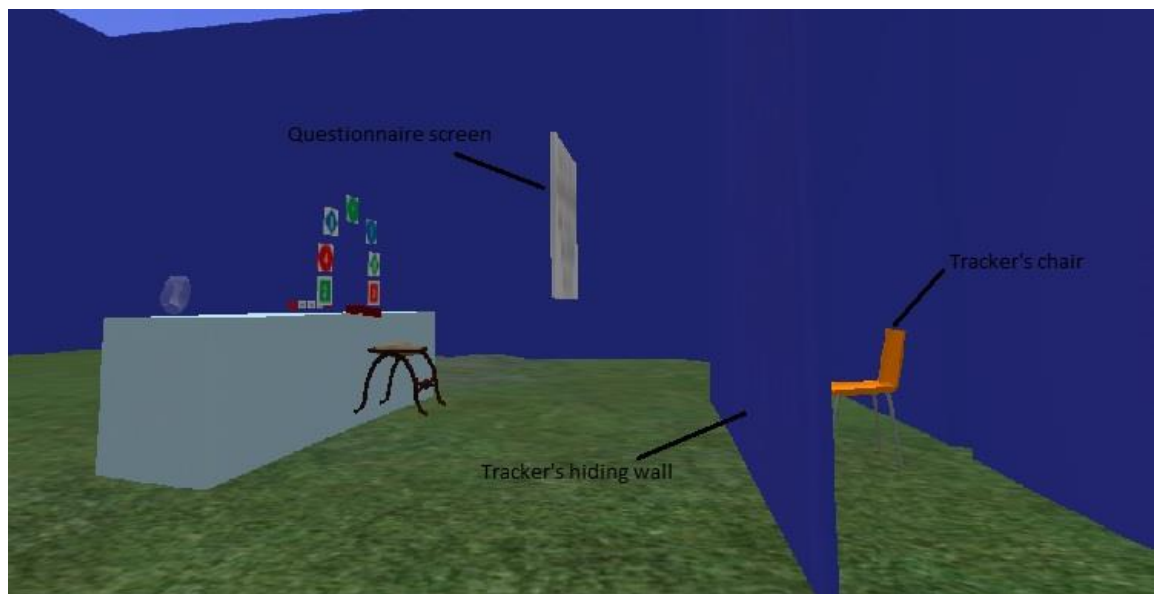


Figure 34: Stimulus and seating set-up

A table and glass screen were built in the same way as the building, and two chairs were placed, one on each side facing each other, one for each avatar (chairs can be purchased ready-made in Second Life; see Figure 34<sup>1</sup>). Both seats contained a script which took control of the user's view once their avatar sat on them. This meant that all participants had the same view of the other avatar, minimising effects due to factors other than experimental manipulation. Although the Tracker program was required to control the elements of the human-human experiment, he was not required to be visibly present within the room; indeed, his presence may have adversely affected the behaviour of the participants. As such, a wall was built within the room, concealing a chair on which Tracker sat, so users were unaware of his presence within the environment (see Figure 35). This chair was also used in the human-agent study to conceal Tracker during presentation of the questionnaire. After the basic and persistent elements of the room had been created, the stimuli could be created.



**Figure 35: Experimental room set-up**

---

<sup>1</sup> This is the set-up for the Human-Agent study; in the Human-Human study, there are a set of Instruction Tiles in front of the Instruction Giver (see Figure 1 in the Human-Human method section).

### 5.5.1 Tasks and Stimuli

The tasks in the original exploratory study involved matching stimuli, completing sequences and identifying mirror images or rotations of shapes. The stimuli involved were dominos, rune-like shapes and geometric shapes. These tasks were inspired by viewing a variety of online IQ tests. The results of the exploratory study indicated a floor effect, with no one completing all tasks correctly. This suggested that the tasks were unduly challenging, given the current paradigm, and that a refined paradigm would require simpler tasks. The main concern was that cognitive load may discourage mutual gaze, since it is known that when cognitive load is high, individuals will avert their gaze from a conversational partner (Glenberg, Schroeder, & Robertson, 1998). A task needed to be selected, therefore, whose difficulty would not interfere with normal patterns of social interaction. It was decided that arithmetic tasks would be most appropriate, since they would not be novel, or require any specialist learning, and they would be universal, rather than language-dependent (i.e. the participant's level of expertise in English would not impact on performance).

One of the fundamental aspects of the human-human and human-agent study was that the verbal instructions involved describing numbers without naming them. This meant that each number being referred to needed a distinguishing feature. As such, it was decided that each number should be set on a background of a coloured shape. Within each trial, only one of each colour-shape combination was visible. Figure 36 shows a selection of stimulus tiles: white numbers set on a background of coloured shapes. White was selected as a text colour as it was deemed to be the most easily visible on all of the coloured backgrounds.

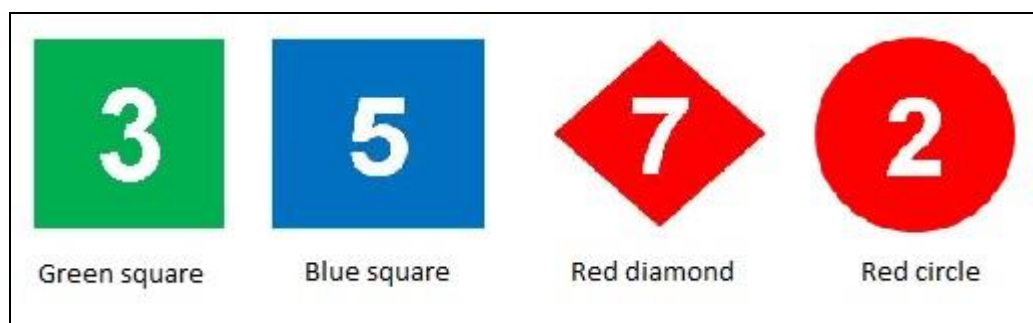


Figure 36: A selection of stimulus tiles

Each of the 30 tasks involved an instruction giver relaying an arithmetic sequence to an instruction follower, referring to numbers on *stimulus tiles*, visible to both participants. The arrangement of the stimulus tiles had to be such that both the instruction giver and instruction follower could see the stimulus tiles. Additionally, the follower needed to be able to see the instruction giver, and where she was looking during each trial. To accommodate these criteria, the stimulus tiles were arranged in an arch formation on a glass screen between the dyad, in such a way as to not occlude either participant's view of the other's avatar (see Figure 34). The instruction giver referred to *instruction tiles* only visible to him or her for the instructions to be relayed to the instruction follower (see Figure 1 in the method section of the Human-Human study). An answer was selected by the instruction follower from three possible *response tiles* in front of them (see Figure 2). Each one of the instruction, stimulus and response tiles was created in the same way as the building; a prim was selected and its position, size and rotation was edited until it was the correct shape and in the correct position. Each group of tiles was then linked together and attached to (linked with) a larger object – the table or the screen between the participants' avatars – to ensure they remained in the correct position, since re-positioning them would be extremely time-consuming. The final property of each tile to be edited was the *texture*.

### **5.5.2 Textures**

In order to present the numbers and provide responses to select, each tile required a texture. These textures were individually created outwith Second Life, saved as jpegs and imported into the environment. Each texture was given a unique name so it could be referred to in the script in order to change the texture on the tiles during the experiment as required. 152 individual textures were created and imported for the human-human study: 81 stimulus tiles, 62 instruction/response tiles, 4 arithmetic operators and 5 background textures.

### **5.5.3 Avatars**

Two avatars were required for the human-human study. Each avatar was dressed and their appearance was edited. The instruction giver's avatar wore a head set with microphone, into which he gave the impression of speaking. He had dark-coloured

eyes, which were clearly visible against his pale skin tone, allowing the instruction follower maximum opportunity to determine the focus of his attention at any one time.

## **5.6 *Experimental Design of Human-Agent Study***

In the human-human study, the tasks were found to be relatively easy, skewing the task performance data. For the human-agent study, the difficulty of the arithmetic tasks was increased to guard against a ceiling effect and provide greater variance within the data. Two of the four blocks involved addition and subtraction tasks, and the other two involved multiplication as well. The sets of tasks were counterbalanced between participants to avoid effects due to type of task. In total, 64 new tasks were created for the human-agent experiment and these were piloted in a small study outside Second Life; the results from this pilot were not subjected to statistical analysis. The majority of the stimulus and response tile textures from the human-human paradigm proved suitable for re-use, although some additional ones were created and imported. Furthermore, an additional 20 questionnaire textures and 5 questionnaire response textures for the human-agent experiment were created.

### **5.6.1 Audio instructions**

Tracker's instructions were presented in the form of a series of individual audio files. Each one consisted of a word or two naming a colour-shape combination (e.g. 'Red Square') or an arithmetic operator. The files were then scripted to play one after the other to form instructions. The introductions presented at the start of each block were also comprised of a series of sound files. 37 individual sound files were recorded in total and imported into Second Life. Each was, again, assigned a unique name. After experimenting with various different voices, one of the female Microsoft Office's Text-to-Speech voices was eventually used. This was mainly because the program was found to be the most straightforward to use.

### **5.6.2 Avatar controller**

To control Tracker, the instruction giver, and the texture changes within the environment, scripts were written, consisting of instructions for moment-by-moment changes within the virtual environment. At the start of every task, the textures on the



instruction tiles and response tiles were set. Next followed instructions for Tracker to look at a stimulus tile and for his mouth to open and close, along with the initiation of the sound file describing the number on that tile. This was repeated for each number being described in the task (four in total), interspersed by sound files and mouth movements for the arithmetic operators. Tracker was also instructed to look at the user for a set amount of time after looking at each tile being described; this varied according to the looking condition, of which there were four (see the Method section of the human-agent chapter for details). The Tracker program was then instructed to wait for a response from the instruction follower, before initiating the script for the next task. An example of an individual log file can be seen in Appendix F: Script 10c.50 (Experiment 2). Additional scripts, such as those to set up the camera view and enable the user to sit on a chair were also written. There were 16 tasks in each block and four blocks, resulting in a total of 64 tasks in the human-agent experiment. Each task was presented under one of the four looking conditions, resulting in a total of over 250 scripts to be written. The total, including the introduction and questionnaire scripts and the additional instruction scripts, resulted in over 300 log files to be written. Each script was checked and each block run through several times to ensure that the textures and corresponding Tracker behaviours – speech, gaze and mouth movements – were presented at the correct times.

### **5.6.3 Questionnaire**

To investigate perceptions of social presence, a suitable questionnaire was needed. Due to the novelty of the research, there was no existing questionnaire that had been validated for use within Second Life, and more specifically with the current paradigm. The most relevant previous study was identified, which included two social presence items within a more general questionnaire (Jin, 2009). It investigated users' perceptions of an avatar's credibility, and the informational value of promotional messages under two modalities of message presentation – text vs. audio – within an Apple store within Second Life. It was decided that the two items included in that study were not sufficient to capture the dimensions of Second Life required for the current study.

After extensive investigation, it was decided instead that an existing questionnaire would be adapted and validated for use within the current paradigm. Witmer and Singer (1998) developed a questionnaire which measured dimensions of presence (not only social presence) within virtual environments. This was deemed to be a suitable starting place for development of our questionnaire, since the questions were related to experience of a virtual environment, so could be easily adapted without having to substantially alter the terminology. For example, Witmer and Singer asked “How natural did your interactions with the environment seem?” and “How much did your experiences in the virtual environment seem consistent with your real-world experiences?” To investigate how similar the interaction with Tracker was to a face-to-face interaction (i.e. how socially present he was, according to our definition), the questions were changed to asked “How natural did your interactions with Tracker seem?” and “How much did your interaction with Tracker seem consistent with your real-world interactions?” Some of the original questionnaire items, such as “How compelling was your sense of objects moving through space?” were deemed to have no social presence-related equivalent, and therefore were not included in our questionnaire.

Since the original questionnaire had been adapted, and the new version was used within a different paradigm, it required validation before any statistical inferences could be made. The new 20-item social presence questionnaire was administered four times during the experiment – once after each block of trials. It was presented in the form of a texture on a questionnaire screen behind Tracker’s seat (see Figure 35). The screen was built in the same way as all of the other objects within the environment, and designed to blend into the environment when not in use. The questionnaire item textures changed in response to button presses on one of 5 response tiles. There were 20 questionnaire textures and 5 questionnaire response textures, all of which were created and imported into Second Life in the same way as previous textures. Tracker ‘left’ the experimental room, and hid behind the hiding wall, allowing the Tracker program to continue without her presence impacting on the responses of the participant. The questionnaire items occupied the screen space previously filled with Tracker.

## 5.7 Troubleshooting

The main issues encountered with the human-human study were related to successful audio and video capture of the experiments. On several occasions it was found that the timings for the audio and video did not match up correctly. For some of these trials, it was possible to edit the audio or video so that they ran correctly together; on other occasions this was not possible, so the trial, along with the rest of that participant's data, had to be discarded. These participants were replaced.

The building and stimuli often required maintenance, which was time-consuming. As previously mentioned, Second Life updates regularly, and this often necessitated alterations to the program controlling Tracker's behaviour and the rest of the experiment.

The most challenging issue in the development of the human-agent paradigm was with *Lipsync*. Tracker, the instruction giver, had to demonstrate behavioural realism in order to appear as a social being within the environment. One of the facets of this behaviour was lipsync – mouth movement corresponding in a believable way to the verbal instructions. It was discovered that, although an avatar is able to lipsync, there was no existing method for controlling a programmed avatar's mouth movements in the same way. Various workarounds were discussed; the initial solution posited being to present Tracker in the form of a video embedded in Second life, giving the impression of him sitting opposite the participant within the environment. This idea was discarded due to difficulty in disguising the fact that Tracker was appearing on a video, rather than being present within the environment. After much investigation, the final solution was to use the existing Second Life 'open mouth' and 'close mouth' commands. Although this was not true lipsyncing, it was discovered that opening and closing the mouth at set intervals during the presentation of the audio instructions was sufficient to give the impression that the lips were synchronised with the words. This was the most time-consuming challenge encountered during the human-agent experimental design.

## 6 Chapter 6: Discussion

Several of the theories and ideas covered in the literature review section of this thesis may help to explain some of the results found in both of the studies. The following section is a discussion of the results from both experiments 1 and 2 as they relate to the literature.

**Table 6: Human-Human results**

	<b>Staring</b>	<b>Not Staring</b>
<b>Mutual Gaze</b>	Less	More
<b>Task Performance</b>	No difference	
<b>MG x Task Performance</b>	No correlation	Positive Correlation
<b>Looking at task-irrelevant</b>	More	Less
<b>Looking at stimulus tiles</b>	Less	More

**Table 7: Human-Agent behavioural results; X indicates no effect**

	<b>Agency</b>	<b>Looking condition (%)</b>	<b>Agency x Looking</b>
<b>Number correct</b>	X	X	X
<b>RT</b>	Avatar < Agent	X	X
<b>MG</b>	X	75>50; 75>25	X
<b>RT X MG</b>	X	Negative correlation in 25%	X
<b>Stimulus tiles</b>	X	75<0; 75<25; 75<50	X
<b>Response tiles</b>	X	X	X
<b>Tracker</b>	X	75>0; 75>25; 75>50	X
<b>Fixation duration</b>	Tracker>Stimulus/Response		

**Table 8: Human-Agent questionnaire results**

<b>Factor score: LookCon</b>	Factor 3: 75%>50% (also 25%,0%;ns)	Social presence perceptions higher in 75% than 50%
<b>RT x Factor score</b>	F3 x RT ( $r = -.328$ )	Higher social presence perceptions related to faster reaction times
<b>RT x Factor score</b>	50% F3 x RT ( $r = -.487$ )	Stronger relationship in 50% condition
<b>MG x Factor score</b>	75% F1xMG ( $r = .447$ )	F1 – social presence behaviours
<b>MG x Factor score</b>	50% F4 x MG ( $r = .461$ )	F4 – immersion / comprehension
<b>DVs as predictors of agency</b>	Overall, only RT made sig contribution	75%: RT and F3 score predict agency

### **6.1 Staring and mutual gaze**

The first aim was to find out if staring at a conversational partner is actually the optimum method for studying mutual gaze: does staring maximise mutual gaze, or is there is an alternative amount of looking that will result in higher levels? Challenging the experimental design of previous studies, we showed that staring may negatively impact on mutual gaze, showing signs of reducing it (Argyle & Dean, 1965; Bailenson et al., 2001). Having a virtual human staring at you does not equate to “engaging them in mutual gaze” (Bailenson et al., 2003); indeed, quite the opposite occurs; staring discourages mutual gaze. One of the reasons for this could have been that which was alluded to by Kendon: “To be subjected to the continual gaze of another is a very unnerving experience, for to be the object of another’s attention is to be vulnerable to him” (Kendon, 1967). If someone is staring at you, discomfort (or perhaps increased arousal) makes you avert your gaze. This theory seems to be supported by the increased looking at task-irrelevant objects in the staring condition in the human-human study.

The hypothesis that unnatural looking behaviour (staring) will dissuade users from engaging in mutual gaze is extended in the second experiment. There was higher uptake of mutual gaze in the 75% condition than in either the 25% or the 50% conditions, but no differences between the latter two looking conditions. This difference between the 75% condition and the other looking conditions is found in several of the dependent variables (which will be discussed later) and will be referred to as the *75% effect*. Although mean mutual gaze uptake was higher in the avatar condition than in the agent condition, the difference was not significant.

## **6.2 Mutual gaze, agency and looking condition: The 75% effect**

No agency effect or agency x looking condition interaction was found for mutual gaze uptake, perhaps in some part due to the size of the error in the data. It can be seen, however, that the mean difference in mutual gaze uptake between the agent and avatar conditions is smaller in the 75% looking condition than in either the 25% or the 50% conditions. This difference, along with other evidence associated with the 75% effect, could be explained by the actions of the IG not being typical of normal human behaviour: by not looking at the instruction tiles before giving instructions, and therefore appearing to hold the instructions for all trials in her head, Tracker behaves differently to how one would expect a human to, and is perhaps perceived as being less than behaviourally realistic in both the avatar and agent condition. Any differences in how the user interacts with an agent vs. an avatar in the 75% condition, therefore, are minimised. None of the differences in mutual gaze uptake between the agent and avatar conditions in any of the looking conditions were significant, however, so although there is a trend towards the user responding to the agent and avatar in a more similar way in the 75% condition than either of the other two, no statistically-supported conclusions can be drawn.

Another explanation for the 75% effect rejects the idea that Tracker's non-typical human behaviour is affecting how the user responds, and relates to the Threshold Model of Social Influence (Blascovich et al., 2002). Within each looking condition, Tracker's behaviour was the same for both of the agency conditions; therefore any

differences in response to her will be due to the instructions that the participant was interacting with either a human or a computer. As long as the level of looking by Tracker is providing a high enough level of behavioural realism, then the threshold of social influence has been reached, therefore whether the user is interacting with a person or a computer should make no difference. If, however, the behavioural realism is not high enough then the level of social presence compensates; users must believe that they are interacting with a person in order for Tracker to exert social influence. If behavioural realism is not sufficient, we would expect to find larger differences between the agent and avatar condition than if it is high enough to reach the threshold without social presence having to compensate; agency only matters if behavioural realism is not high enough. The smaller difference between agent and avatar in the 75% condition supports the idea that, in looking at the user during 75% of the trial, the realism of Tracker's behaviour has potentially reached the threshold of social influence, therefore agency does not matter. It is not that the behaviour is far from being typical human looking behaviour, but that it is the *closest to* typical (compared with the other looking conditions).

This theory fits with Wang and Gratch's (2010) report that speakers look at listeners, on average, during 65% of an interaction, suggesting that the threshold may be reached somewhere near that point. One would perhaps expect an interlocutor not to be looking away at something you can't see when she is helping you complete a task, and the fact that she is retaining all of the information in her head is irrelevant. We are used to people reading from autocues, on the news for example. Indeed, Krämer (2012) maintains that "Interactions with computers and artificial agents have already been likened to the interaction and relationship we have with media persona (e.g. an anchorman/woman in TV news)" (p.216). Perhaps not looking at the instructions is not surprising at all. After all, the participant did not know that the IG was reading the instructions when she looked down, only that she was not looking at them (the user). We are, however, dealing with trend here, rather than statistical differences, so other results relating to the 75% effect need to be examined to establish which explanation of this trend is correct. What is clear is that there are two contrasting potential explanations for the 75% effect and both will be discussed in relation to the remaining findings.

Another result showing the 75% effect is the additional looking at Tracker (and longer fixation duration), and less looking at stimulus tiles, compared with the lower three looking conditions. The increased mutual gaze uptake in the 75% is likely be a result of the overall increased looking at Tracker. Again, this could be a result of curiosity, an artefact of attending more to the less than typical human behaviour in the 75% looking condition. Conversely, Tracker may be more engaging when looking 75% of the time, since there she displays *more* behavioural realism, therefore she is attracting more attention. Since agency is held constant when comparing looking conditions, behaviour (and behavioural realism) is the only variable which can affect social influence. As previously discussed, the difference between agent and avatar in the 75% condition could provide clues as to what is going on.

According to Blascovich's social influence model, agency matters when the behavioural realism is not high enough to meet the threshold, so low levels (of behavioural realism) would result in larger differences between the agency conditions. This is assuming that looking at Tracker is a socially-driven response during an interaction, resulting from her exerting social influence. As can be seen from Figure 20, there is little difference between the agency conditions in the amount of looking at Tracker in the 75% condition (compared with 50% and 25%), meaning that the threshold has potentially been met. There is also, however, little difference between agent and avatar in the 0% condition. One would assume that not looking at a conversational partner at all during an interaction could be deemed to be non-typical behaviour, which could result in the user treating agent and avatar similarly (as previously discussed). Again, we are only looking at trends, although they may help build a picture of how the user perceives Tracker's looking behaviour. Because users are not being stared at (as in experiment 1) their gaze is not averted for social reasons. It is unlikely that this is directed mutual gaze, i.e. deliberate, task-related mutual gaze, since there is no task benefit in the additional looking by Tracker in the 75% condition. There is no detriment to task performance, however, so this additional looking at Tracker and less looking at Stimulus tiles is not hindering task performance. It would follow, therefore, that the additional looking at Tracker is socially-driven (and therefore a result of Tracker exerting social influence), but it is



still not obvious what, specifically, is driving it. Again, we see a difference between 75% and the other three conditions (but no difference between 0%, 25% and 50%), demonstrating that there is something different about this higher amount of looking by Tracker. It is still unclear if this is a result of higher or lower levels of behavioural realism.

### ***6.3 The function of increased looking at the IG: Addressing a critique***

One of the aims of methodology design in experiment 2 was to address a critique of the human-human study – that increased looking at Tracker may have been an artefact of extra visual information presented in the not-staring condition. If the extra looking at Tracker (and uptake of mutual gaze) in experiment 1 was caused simply by the holding of visual information that helped complete the task, then we would expect to find that Tracker attracted the same amount of looking in all of the looking conditions in experiment 2. We did find that for the 0%, 25% and 50% conditions, but not for the 75% condition. It could well be, therefore, that the amount of looking at Tracker in the three lower looking conditions is, indeed, task-driven since it is the same amount whether Tracker does not look at the user at all – i.e. the baseline 0% condition – as when she looks at the user 25% and 50% of the time. The extra looking in the 75% condition, however, is socially driven, because the user wants to engage in eye contact. The curiosity of Tracker's behaviour or social influence she exerts due to levels of behavioural realism attracts attention, resulting in increased mutual gaze, since the user does look more at Tracker although no extra task-related visual information is being presented in this condition. It can be concluded, therefore, as expected in the experiment 1 methodology, that the extra looking at stimulus tiles in experiment 1 was, indeed, redundant.

Agency did not have any effect on the amount of time spent looking at each interest area type, and there was no interaction between agency and looking condition. Since there was a higher uptake of mutual gaze in the 75% looking condition, yet no main looking condition effect on task performance (i.e. task performance was not

significantly better in the 75% condition), how was the amount of mutual gaze related to task performance in this condition, or indeed in any of the other looking conditions?

#### **6.4 Task performance and social facilitation**

The main task performance finding was that tasks were completed significantly faster in the avatar condition than in the agent condition. One explanation for this is that users were unwilling to keep a person waiting, whereas they did not feel that they needed to hurry for a computer. If this was the case, however, we would expect to see differences in the number correct: faster reaction times to the detriment of accuracy. There was no effect of agency on accuracy scores. Another explanation is that it is evidence of social facilitation (Zajonc, 1965). If users are treating Tracker as a human, and social facilitation is occurring, we would expect to find a difference between the task performance in the agent and avatar conditions, with users performing better on the task when they believe a human is present than when they think they are interacting with a computer. It was, indeed, found that users completed the tasks significantly faster when they believed they were interacting with a human, not a computer, providing evidence of social facilitation in the avatar but not the agent condition. Furthermore, it is evidence in favour of the Threshold Model of Social Influence (Blascovich et al., 2002). When averaging over looking conditions (some of which we believe to be behaviourally unrealistic), the fact that they had been told that they were interacting with a human was enough to reach the threshold, enabling the avatar to have social influence over them (but not the computer-controlled agent), and resulting in social facilitation. A number of participants reported that they did not believe the agency manipulation, but the data tells a different story. Coupled with this, there was no difference in self-report social presence measures (factor 3 score). This suggests that this behavioural measure (RT) is more sensitive to variations in agency of an interlocutor than self-report measures.

To investigate further the 75% effect, the reaction times were split by agency and the different levels of looking condition were compared. It was found that reaction times

were faster within the avatar group in the 75% condition than the 50%, and there was a trend towards quicker RTs in 75% compared with 25% and 0%. The same pattern was not found in the agent condition. When users believe they are interacting with a human, they will respond faster in this anomalous 75% condition, but not when they are interacting with a computer. The theory that users did not want to keep their human interlocutor waiting is weakened here; if that was the case then amount of looking by Tracker should make no difference. How does this tie in with the idea that the 75% condition shows non-typical human like behaviour? It seems counter-intuitive: the user thinks the IG is acting behaviourally unrealistically in the 75% condition, yet will still perform faster in order not to keep her waiting. Instead the results provide more support for the theory that at 75% looking Tracker exerts more social influence, resulting in more social facilitation, but only when interacting with a human. Since users are interacting with a human, and behavioural realism is higher in the 75% condition (compared with the other looking conditions), the threshold has been reached and surpassed, resulting in increased social influence and more social facilitation.

All of the 75% effect results lead us to ask what it is that is special about looking at one's interlocutor 75% of the time, as opposed to 50%, 25% or 0% of the time. Any looking condition effects don't appear to be linear – there is a lack of observable increase or decrease as the amount of looking at user increases, and the majority of looking condition effects can be seen in the 75% looking condition. The essential difference between the four looking conditions is that in all but the 75% condition, Tracker looks at the instruction tiles before communicating the task. This may be enough to provoke strong behavioural differences in the user, and result in the 75% effect.

Mutual gaze was found to facilitate task performance in the human-human experiment, supporting evidence from face-to-face and video mediated interactions involving different types of cognitive tasks (Fry & Smith, 1975; Fullwood & Doherty-Sneddon, 2006). It is also evidence of social facilitation effects during CMC, via human-controlled avatars. As with Blascovich et al (2002) and Zambaka et

al (2007), there was some evidence of a ceiling effect, in that the data was positively skewed, although in our case we did find a significant effect. One would expect, however, to find social facilitation effects in both the staring and not-staring conditions; since users knew they were interacting with another human, the threshold of social influence should have been reached by agency alone, despite the staring potentially lacking behavioural realism. The fact that the relationship between task performance and mutual gaze in the staring condition failed to reach significance suggests that something else was happening to inhibit that performance, namely Tracker staring at the user. Could the staring be such non-typical human behaviour (and therefore behaviourally unrealistic), meaning that the followers are treating the IG more as an agent than an avatar? This could be evidence for the opposite to Ethopoeia: despite knowing they are interacting with another person, the non-standard human behaviour (staring) is encouraging them to perceive their interlocutor more as an agent, resulting in no facilitation effects. Or it could relate to some aspect of the threshold model of social influence. Zambaka and colleagues already found some evidence that interaction with an avatar can result in social inhibition effects, where none exist during agent interactions. It would seem that this result extends this by finding a lack of social facilitation effects when the avatar displays non-typical human behaviour (staring), leading the user to treat it more like a computer. In terms of the threshold model of social influence, is it possible that the behavioural realism was so low in the staring condition that the threshold was not reached and social facilitation did not occur?

The experiment 1 findings were replicated in experiment 2, with a significant negative relationship between uptake of mutual gaze and reaction time, but only in the 25% condition: the higher the uptake of mutual gaze, the shorter the RT. This finding echoes that from the human-human experiment. When the experiment 2 results from the 25% condition were split by agency, a significant negative relationship was found between uptake of mutual gaze and reaction time in the avatar condition, but not in the agent condition. When the user believes they are interacting with a human, increased mutual gaze may result in better task performance, but only if the overall opportunities for mutual gaze are 25% of the interaction. When they

believe that they are interacting with a computer, however, this is not the case. How this result fits with the threshold model of social influence and theory of social facilitation/inhibition is potentially complex. A clear difference in behaviour can be seen in the agent and avatar conditions, with users who believe they are interacting with a human performing better with increased mutual gaze. However, because there is a lack of significant relationship between mutual gaze and task performance in the 75% condition (even when split by agency), it appears that additional looking at Tracker in this condition, and the resulting increased eye contact, have no task-related benefit, strengthening the idea that the extra looking is socially-driven. As previously discussed, reaction times were, on average, faster within the avatar group in the 75% condition (than the other looking conditions) but the same pattern was not found in the agent condition. Users in the 75% condition looked more at Tracker and engaged in more mutual gaze, but this extra looking did not facilitate task performance, even when split by agency.

In a number of cases, during a debriefing at the end of the experiment participants in the avatar condition claimed not to have believed the agency manipulation. This could be to save face (to avoid admitting to being tricked by the manipulation), but if it is true then there should be no agency effects. An agency effect on any of the behavioural outcomes could suggest that these are more sensitive measures of a user's social perceptions of an interlocutor than self-report measures. There was an overall agency effect on task performance, in terms of reaction time: users were quicker to respond in the avatar condition than in the agent condition. This supports the theory of social facilitation (or that they were less inclined to keep a person waiting than a computer). The theory of social facilitation says that if the tasks are well-learned tasks, then the dominant response (in the presence of others) will be facilitated, i.e. successful completion of the task. As can be seen in Figure 14, the dominant response was correct (i.e. the mean score for all conditions was more than 5/10, with the 95% of the scores not falling below 6/10). It follows, therefore, that task performance would be facilitated by the presence of another person, and less so by the presence of a computer, which wields less social influence. As in the human-human study, there was no main effect of looking condition on task performance.

This means that any differences in task performance are likely to be more complex than simply varying the amount that one conversational partner looks at another. To gain a fuller understanding of what was going on, the questionnaire data was analysed.

## **6.5 Social presence**

Following a principle components analysis, an adequate 4-factor structure was discovered for the responses to the 20 items making up the questionnaire. This echoed the structure of the questionnaire on which the questions were based (Witmer & Singer, 1998). The factor of most interest here was factor 3 – social interaction/perceptions. This was the one that best encapsulated the notion of social presence. It was anticipated that, if the agency manipulation had been believed, then an agency effect would emerge in the factor 3 factor scores. No such effect was found. It seems that, despite being designed specifically to measure social presence, our questionnaire was not sensitive enough to detect variances in agency. We have already discovered that there is a reaction time effect of agency, and that there was an effect of looking condition on RT in only the avatar condition and not agent, and therefore this adds weight to the idea that behavioural measures (i.e. RT) are more sensitive predictors of how socially present an interlocutor is. It could also relate to the idea that some users maintained that they did not believe the agency manipulation; it would make sense that behavioural measures were needed to detect the differences that self-report measures could not fully capture.

Previous research during interactions with an agent/avatar within an IVE found that gaze behaviour was related to social presence ratings; social presence ratings were higher in a high gaze condition (eyes fixed on the participant and head turning to follow them around the room) than in a low gaze condition (Bailenson et al., 2003). We would expect, therefore, to find an effect of looking condition on social presence ratings. A main effect of looking condition was, indeed, found on factor 3 score. It was found that participants scored significantly higher on factor 3 in the 75% looking condition than in the 50% looking condition. No other significant differences were

found. This suggests that users perceive their interlocutor to be more socially present in the 75% condition than in the 50% condition. This echoes the finding that users were faster in the 75% looking condition than the 50% condition in the avatar group as they did not want to keep their human conversational partner waiting, or because higher levels of behavioural realism led to more social facilitation: Tracker was more socially present in the 75% condition than in 50%. The differences between 75% and either 0% or 25% were non-significant. It can be seen in Figure 25 that the mean factor 3 score was lower in the 50% condition than any of the other three, and the difference between 50% and 25% was non-significant (see Table 5). It could be that the behavioural realism in the 50% condition was lower than that in 75% and 25%, leading to users reporting Tracker to be less socially present in this condition. The size of effect for the difference between factor 3 score in the 75% and 50% looking conditions, however, was not large ( $d = .49$ ; a  $d$  of  $>.5$  signifies a medium effect (Cohen, 1977)). That, coupled with the fact that the differences between the 75% and either 0% or 25% conditions were not significant (where we have previously found task performance and eye movement effects) suggests that the self-report measures aren't necessarily detecting all of the differences due to social presence.

Factor 3 was the only one that yielded a significant relationship between task performance and factor scores. A significant negative correlation between factor 3 scores and overall reaction time from end of instructions was found, indicating that as social presence perceptions increase, participants are quicker at completing the tasks. Furthermore, this negative relationship was found to show a tendency towards significance in the avatar condition, but was non-significant in the agent condition. This finding tallies with the agency effect on reaction time: participants were quicker to complete tasks when they believed they were interacting with a human than when they had been told their interlocutor was a computer. This could, again, be due to an unwillingness to keep a person waiting but that a computer would not mind waiting for the tasks to be completed, or could be due to social facilitation. The social facilitation explanation and idea that users did not want to keep a human waiting are both evidence that Tracker is exerting social influence over the user under certain conditions, each emphasising a different aspect of social influence.

The majority of effects were found within the 75% condition and for factor 3. There were, however, some other relationships that could not be related to the 75% effect. A significant negative relationship was found between reaction time and factor 3 score in the 50% looking condition. When split by agency this pattern was found in the agent condition but not the avatar condition. When users are being looked at 50% of the time, a faster reaction time is associated with a higher social presence rating. This seems to add support to the social facilitation theory. What is less easy to explain, however, is that higher social presence ratings were associated with faster reaction times in the agent condition but not in the avatar condition.

When looking at the relationship between mutual gaze uptake and factor score in the 75% looking condition, a positive correlation was found between factor 1 score and mutual gaze uptake. When split by agency, the correlation showed a trend towards significance for agent and was non-significant for avatar. This suggests that in the 75% looking condition, more mutual gaze is associated with higher behavioural/interface ratings, but when split by agency, the data suggests that this pattern can only be seen when the user thinks they are interacting with a computer, and not a human. The positive relationship between factor 3 score and mutual gaze uptake in the 75% looking condition was found to be non-significant.

Within the 50% looking condition, a significant positive relationship was found between factor 4 score and proportion of mutual gaze uptake. When split by agency, the correlation was significant for avatar but not for agent. Factor 4 related to techniques learnt during the interaction as well as the impact of control devices. In the 50% looking condition, therefore, more mutual gaze is related to higher perceptions of techniques learnt during the interaction and of the control devices. Since only 2 items loaded strongly onto this factor, however, it is difficult to interpret these results with confidence.

The final analysis looked at reaction time, mutual gaze and factor score as predictors of the outcome variable agency. It was found that overall reaction time was the only significant predictor of whether the participant was in the agent or avatar condition. When split by looking condition, in the 50% condition reaction time, again, was the



only predictor to make a significant contribution to the outcome agency. In the 75% looking condition, reaction time made a significant contribution to the outcome, and factor 3 score made a marginally significant contribution. It is interesting that the task performance measures detect the differences between the agent and avatar conditions better than the self-report social presence questionnaire (i.e. factor 3 score). The only place where social presence measures detected agency differences was in the 75% condition, and even then the effect was marginal. This adds further weight to the idea that behavioural measures (i.e. reaction time) could help to identify differences between perceptions of an interlocutor. One could argue that participants did not believe the agency manipulation, but the differences in task performance mean that they were clearly responding in a different way since the agency instructions were the only difference between the two conditions. It could be that they were claiming not to have been taken in by the instructions in order to save face. Investigation into the relationship between whether they believed the agency manipulation and the other variables, especially RT and social presence measures, could assist in the overall understanding of the model.

Overall, the model confirms the previous RT x agency effect, and shows that reaction time is better predictor of agency than any of the other variables. Most interestingly, RT is able to predict agency better than questionnaire responses, which is counter-intuitive since the questionnaire was designed specifically to detect variations in social presence. The contribution of factor 3 score was marginally significant in the 75% looking condition, which suggests that under some circumstances a combination of behavioural measures (reaction time) and subjective measures (social presence questionnaire responses) are best able to predict the outcome agency.

The final chapter in this thesis revisits the research questions presented at the end of Chapter 2, and explores avenues of future research, motivated by the two studies described in this thesis.

## **7 Chapter 7: Conclusions and future work**

### ***7.1 Research questions revisited***

The answers to the research questions presented at the end of chapter one can now be summarized as follows:

1. What is the optimum amount of looking by one individual at another for maximising mutual gaze, task performance and social presence?
  - Having one person stare continuously at another may not be an appropriate method for studying mutual gaze; staring discourages mutual gaze, therefore such an experimental method would negatively impact on outcome measure.
  - The optimum amount of looking by one conversational partner at another varies from situation to situation and depends on the purpose of the interaction. 75% looking produced the highest mutual gaze uptake and showed evidence of increased social presence perceptions. In 25% looking, however, mutual gaze was found to facilitate task performance, but only when not being stared at.
2. What is the relationship between purported agency (manipulated social presence), mutual gaze, task performance and perceived social presence? Can behavioural measures in this paradigm (gaze/task performance) identify variations in actual social presence which may go undetected by a self-report questionnaire?
  - This research shows the behavioural measure, reaction time, to predict actual/manipulated social presence. Self-report measures alone were unable to detect differences in actual social presence, although there was evidence to suggest that, under certain conditions, a combination of behavioural (RT) and self-report measures may be beneficial.
3. What is the function of gaze, specifically mutual gaze, during task-based interactions within Second Life?

- Mutual gaze can facilitate task performance in a task-based interaction within Second Life when being looked at during 25% of the interaction.
  - This does not hold true, however, if you are being stared at.
- Additional mutual gaze in the 75% condition was not accompanied by increased task performance, however, suggesting that it was socially, rather than task-driven.

Mutual gaze can be task or socially driven, and this can depend on additional mediating behaviours, such as the amount of looking by one conversational partner at another.

## **7.2 Future work**

This thesis aimed to examine the relationship between mutual gaze, task performance and social presence during mediated interactions in Second Life. The discovery that task performance can be facilitated by increasing mutual gaze has implications for many areas of life, from business meetings to retail to pedagogy, including virtual teaching agents, and perhaps even face-to-face teaching. Mutual gaze matters during mediated social interaction. However, staring at someone neither maximises the mutual gaze nor assists task performance. Furthermore, mutual gaze does not always maximise task performance: there is an optimum level of speaker/instruction giver looking at which mutual gaze and task performance are positively correlated, and this may not be the same level that results in the most mutual gaze.

As suggested in experiment 1, gaze aversion when being stared at would seem to be driven by social factors (discomfort at being stared at drives the user to look anywhere other than at their interlocutor's eyes), whereas looks away from Tracker when not being stared at were task-related (purposeful looking away in order to look at task-related objects and gain information about the task). Future work could investigate this further by examining the sequence of looking: where does the user look after glances at Tracker, and how does this vary between gaze conditions? This could assist in furthering the understanding of how people respond to specific gaze patterns.

The 75% effect seems to provide evidence in favour of the Threshold Model of Social Influence (Blascovich et al., 2002); by the time a speaker is looking at his listener during 75% of the interaction, a threshold has been reached, meaning that Tracker exerts social influence over the follower. This results in the maximising of mutual gaze and increased social perceptions of the instruction giver. We have no way of knowing, however, what the behaviours would be if the IG looked at the user, say, 35% of the time, or 60%, since we only have 4 reference points and no simple emerging pattern of increase or decrease in any of the DVs as the amount of IG looking varies. It may be of special interest to find out what happens between the 75% and 100% points. Experiment 1 found that 100% looking did not maximise mutual gaze or task performance, but Experiment 2 suggests there are some benefits to being looked at during 75% of the interaction (compared with the other levels of IG looking), in particular increased mutual gaze uptake and faster reaction times when the user believes they are interacting with a human. It would therefore be interesting to establish the pattern between 75% and 100%, with levels of IG looking set at between 80% and 95%. Including points between 50% and 75% could also help us investigate if there really is a behavioural realism threshold, and if, as Wang and Gratch (2010) suggest, this lies at around 65%. If so, we would expect to find, in the 65% condition, similar social presence perceptions and mutual gaze uptake to those found in the 75% looking condition, potentially even stronger effects. It was impractical to include any more looking conditions on this occasion, but further work could expand on these results to include different amounts of IG looking. This may provide a fuller understanding of, for example, the 75% effect (or perhaps a 65% effect), or any other effects relating to different amounts that the speaker looks at the follower.

Based on the results from this thesis, how does one go about designing the looking behaviour of a computer-controlled agent or human-controlled avatar? The answer seems to be: It depends on the function of that interaction. If the purpose is to maximise social presence between the pair then 75% looking would be appropriate. In a counselling situation, for example (Yuen et al., 2013), it would make sense to maximise the sense of social presence between a dyad to make it as similar to a face-to-face interaction as possible. This would be especially appropriate since mutual

gaze is also maximised at this level of looking. As previously discussed, mutual gaze has various functions, including signalling that a conversational channel is open, assisting in information seeking and establishment and recognition of social relationships (Argyle & Dean, 1965). These would all seem to be important factors in a counselling situation. If, however, the user is being assisted in completion of a task, and mutual gaze is necessary for completion of the task (for example if shared attention needs to be utilised in order to be successful), then the current evidence suggests that 25% looking would seem to be appropriate, as at this level of looking, task performance is facilitated by mutual gaze.

It is important here to note the *type* of task involved in the interaction; the current studies looked at arithmetic tasks, whereas other cognitive tasks (such as linguistic or memory-based, for example) may require different levels of looking in order to maximise performance. Further investigation should be made to establish optimum looking behaviours for maximising task performance when alternative types of cognitive tasks are involved. A further study would use the same paradigm within Second Life, with the same IVs and DVs, but using a linguistic task, or a memory task. A memory task would be directly imported into the existing paradigm, with stimulus tiles being presented and the IG instructing the IF to remember a sequence of numbers, referring to them by their shapes/colours (as in the experiments covered in this thesis). The IF would then select the correct response from the response tiles, as before. A linguistic task would require more of a re-design of the paradigm, perhaps involving a joint task requiring communication and collaboration for successful completion. It could use the existing arrangement of avatars/agents seated facing each other, with the glass screen between them on which could be presented task-related stimuli, such as words.

This paradigm allows the manipulation of the IG's voice; it may be of interest to manipulate the voice to become more mechanical (less behaviourally realistic) to establish if that has an effect on social perceptions of, and behaviour towards the virtual human. Other manipulations which may have an effect on social presence perceptions are method of presentation of instructions (verbal, text) or visibility of the IG (visible, invisible). It is anticipated that different combinations of these

independent variables would elicit different social presence evaluations. These manipulations could inform further on the most appropriate methods for presentation of information in interactions for different purposes.

Is it important that, during a mediated interaction, a user is conversing with another human (an avatar), or will a computer-controlled agent suffice? Again, it depends on what the dyad hopes to achieve from the interaction. If a well-learned task is the aim, such as arithmetic (or perhaps even online shopping), then it is important that the user interacts with (or at least believes they are interacting with) another human. This thesis does not cover the area of social inhibition, since the tasks involved were well-learned (we can assume that the participants will have learnt arithmetic). It would be interesting to establish whether only avatars will evoke social inhibition effects (as found by (Blascovich et al., 2002), or whether both agents and avatars can (as found by Zambaka et al., (2007) within the current paradigm. A future study would manipulate the tasks to ensure that some were difficult (or novel), enabling conclusions about whether social inhibition effects can be found during interactions with both humans and computers, as found by Blascovich. If inhibition effects were discovered in interactions with avatars, but not agents, then it would follow that when tasks are novel or difficult, interacting with a computer would be the desirable option, since there are not many situations where the aim would be to inhibit performance. Arousal is higher when interacting with another human, leading to facilitation of well-learned or easy tasks, and inhibition of novel or difficult tasks.

There is scope for analogous face-to-face human-human experiments, to further test the relationship between human-avatar interaction and interaction in the real world. This would provide a comparison to the mediated interaction (using agents and avatars). The set-up would be similar to that in the human-human study, but would involve the participants facing each other in the same room with a glass screen between them. The screen would have the stimulus tiles projected onto it for both participants to view, and the IF would respond using a keyboard situated in front of them (as it was within the virtual environment). This could enable us to extend the social presence findings; would we find more social facilitation in the face-to-face (i.e. a maximally socially present) set-up? And how would it affect mutual gaze? As

discussed in the literature review, there is a difference between how people view faces in a real-world interaction and during video viewing of the same interaction. A face-to-face recreation of our study would enable us to find out how similar our interaction with social cues from agents and avatars is to one in the real world. There would, of course, be limitations: experimental control cannot be as high (although this could be increased by using a confederate as an instruction giver), it will not be so easily replicated, and the results could not be directly compared with the computer-mediated ones, due to the differences in set-up. Any findings, however, could assist in not only the understanding of the validity of the use of mediated interactions in the study of social behaviour, but also in furthering the overall understanding of eye contact during social interaction.

Overall, as predicted, we found evidence that behavioural measures were more sensitive at detecting differences in manipulated social presence (agency) than the self-report questionnaire alone. Reaction time was found to better predict manipulated social presence than questionnaire responses. Contrary to predictions, however, our data lent little support to the idea of using mutual gaze as a predictor of social presence. Although there is precedent for the claim of using behavioural measures to predict social presence (Bailenson et al., 2004), it should be noted that the factor structure for the questionnaire that we designed and administered, although adequate, could benefit from some improvements. Future work would re-examine the factor structure of the questionnaire and investigate the effect of removing some of the items or factors to see if the structure could be improved. Factor 4, for example, had only two items loaded strongly onto it, which suggests that the effect of removing them, along with the factor itself, may strengthen the overall questionnaire. The questionnaire would then be re-tested and validated before use in future work within this paradigm. Although a suitable questionnaire needed to be designed specifically for use in this research, the re-testing and re-validation was out-with the scope of the current project. It is anticipated that work using an improved questionnaire, although perhaps more sensitive to differences in manipulated social presence (agency), would still benefit from the addition of behavioural measures, such as reaction time or eye movements, in order to gain a fuller perspective of participants' social perceptions of their virtual interlocutor.

One aspect which has not been addressed in this thesis is that of gender. There is some evidence that gender may impact on behaviour towards and perceptions of an interlocutor. In terms of perceptions, the Vuilleumier et al. (2005) study covered in section 2.3 investigated the effects of gaze perception on face processing. They found stronger effects (gender judgement and post-test recognition) when the face being viewed was of the opposite gender to the observer. Kulms et al. (2011) found that female agents were evaluated more positively and that gaze avoidance was evaluated negatively (see section 2.4). The final study relating to gender and social perceptions that has been discussed in this thesis is that of Bente et al. (2007). They discovered that females were more sensitive to variations in their male partners' gaze than vice versa. The paradigms in these three studies are diverse – one involved participants viewing photographs, one had dyads communicating via avatars and participants interacted with a rapport agent in the other.

The other area where gender has been found to have an effect was on behaviour towards an interlocutor. Bente et al. (1998) found gender differences in non-verbal activity (body movement and gaze). Mulac et al. (1987) found gender differences in same- and mixed-sex dyads on gaze and mutual gaze during problem solving dyadic interactions. Gender can also impact on cognitive performance: men's cognitive function can be temporarily impaired following interaction with women (Karremans, Verwijmeren, Pronk, & Reitsma, 2009) There is even evidence to suggest that men's cognitive performance can be impaired at the *anticipation* of interaction with a woman (Nauts, Metzmacher, Verwijmeren, Rommeswinkel, & Karremans, 2012)

It would be interesting to discover if there would be any effects of gender on perceptions of, or behaviour towards an interlocutor within the paradigm described within this thesis. A future study would repeat the human-agent experiment, using congruent and mismatched gender pairs. This may uncover differences between how each gender uses the eye movements of a gender-congruent and mismatched avatar/agent. It would also discover if agency is affected by gender differences as well as mutual gaze and task performance. Finally, social presence measures would also be included within the paradigm. In using virtual environments to investigate gender differences, it is easier to minimise experimental artefacts (in comparison



with using videos of real humans) and match male-female pairs of agents/avatars more closely; the appearance of the virtual humans can be carefully manipulated to match size / salience of features etc. between manipulated male-female matched pair virtual humans.

The relationship between mutual gaze, task performance and social presence in a mediated environment is a complex one. Maintaining the quality of a mediated interaction, and retaining as many real-world social cues as possible, requires careful manipulation of an avatar or agent's non-verbal behaviour. The purpose of the proposed interaction is important, since specific tasks require different levels of non-verbal behaviours for successful completion. It is anticipated that this thesis will provide a starting point for further investigation into eye movements during mediated social interaction. Furthermore, it will also assist creators of virtual humans in designing the most appropriate behaviour for successful human-computer and mediated social interaction.

## 8 References

- Abeelee, M. V., Roe, K., & Pandelaere, M. (2007). *Construct validation of the concepts social presence, emotional presence and connectedness and an application of Zajonc's social facilitation theory to social presence research*. Paper presented at the Proceedings of 10th international workshop on presence.
- Andrist, S., Mutlu, B., & Gleicher, M. (2010, April 10–15, 2010). *Conversational Gaze Aversion for Virtual Agents*. Paper presented at the CHI 2010, Atlanta, Georgia, USA.
- Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, 2012, 13.
- Argyle, M. (1970). Eye-contact and distance: a reply to Stephenson and Rutter. *British Journal of Psychology*, 61(3), 395.
- Argyle, M. (2013). *Bodily communication*: Routledge.
- Argyle, M., & Dean, J. (1965 -a). Eye-Contact, Distance and Affiliation. *Sociometry*, 28(3 ), 289-304.
- Bailenson, J. N., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., & Blascovich, J. (2004b). *Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments*. Paper presented at the 7th Annual International Workshop on Presence Valencia, Spain.
- Bailenson, J. N., Beall, A. C., & Blascovich, J. (2002). Gaze and task performance in shared virtual environments. *The journal of visualization and computer animation*, 13(5), 313-320.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium Theory revisited: Mutual gaze and personal space in virtual environments. *Presence* 10(6), 583-598.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7 ), 819-833.
- Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. (Vol. 14, pp. 379–393). *Presence: Teleoperators and Virtual Environments*.
- Bente, G., Donaghy, W. C., & Suwelack, D. (1998). Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. *Journal of Nonverbal Behavior*, 22(1), 31-58.
- Bente, G., Eschenburg, F., & Aelker, L. (2007). *Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication*. Paper presented at the Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain.

- Bergum, B., & Lehr, D. (1963). Effects of authoritarianism on vigilance performance. *J Appl Psychol*, 47(1), 75-77.
- Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication*, 3(2), 0-0.
- Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence*, 12(5), 456-480.
- Blagrove, E., Foulsham, T., Watson, D., Payne, L., & Kingstone, A. (2012). Fishing for faces: Looking behaviour inside and outside the lab. *Journal of Vision*, 12(9), 36-36.
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2), 103-124.
- Buchanan, H., Markson, L., Bertrand, E., Greaves, S., Parmar, R., & Paterson, K. B. (2014). Effects of social gaze on visual-spatial imagination. *Frontiers in psychology*, 5.
- Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communications Monographs*, 54(1), 19-41.
- Carletta, J., Evert, S., Heid, U., & Kilgour, J. (2005). The NITE XML toolkit: data model and query language. *Language resources and evaluation*, 39(4), 313-334.
- Cassell, J., Sullivan, J., & Prevost, S. e. (1999). *Embodied Conversational Agents*: MIT Press.
- Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5), 519-538.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev: Lawrence Erlbaum Associates, Inc.
- Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental psychology*, 34(1), 28.
- Dalzel-Job, S., Nicol, C., & Oberlander, J. (2008). *Comparing behavioural and self-report measures of engagement with an embodied conversational agent: A first report on eye tracking in Second Life*. Paper presented at the The 2008 Symposium on Eye Tracking Research & Applications, Savannah, Georgia.
- Dalzel-Job, S., Oberlander, J., & Smith, T. J. (2011). *Don't Look Now: The relationship between mutual gaze, task performance and staring in Second Life*. Paper presented at the *The 33<sup>rd</sup> Annual Conference of the Cognitive Science*, 20-23 July 2011, Boston, Massachusetts, USA.
- Dalzel-Job, S., Papananagiotou, P., Fourneron, N., Vattoly, B., & Tajwer, A. (2006). *Examining the efficacy of "genie" the tour guide*. Unpublished Edinburgh University MSc project.
- Dashiell, J. (1930). An experimental analysis of some group effects. *J Abnorm Soc Psychol*, 25(2), 190-199.
- de Greef, P., & IJsselsteijn, W. (2000). Social presence in the PhotoShare tele-application. *Proceedings of PRESENCE*, 27-28.
- De Kort, Y. A. W., IJsselsteijn, W. A., & Poels, K. (2007, October). *Digital Games as Social Presence Technology: Development of the social presence in*

- gaming questionnaire*. . Paper presented at the The 10th International Workshop on Presence, Barcelona, Spain.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision research*, 51(17), 1920-1931.
- Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PloS one*, 8(1), e53286.
- Fry, R., & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. *Journal of Social Psychology*, 96(1), 145-146.
- Fullwood, C., & Doherty-Sneddon, G. (2006). Effect of gazing at the camera during a video link on recall. . *Applied Ergonomics* 37, 167-175.
- Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & cognition*, 26(4), 651-658.
- Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International journal of educational telecommunications*, 1(2), 147-166.
- Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education*, 11(3), 8-26.
- Harms, C., & Biocca, F. (2004). Internal Consistency and Reliability of the Networked Minds Social Presence Measure. *EXPLORING THE SENSE OF PRESENCE*, 246.
- Hiltz, S. R. (1994). *The virtual classroom: Learning without limits via computer networks*: Intellect Books.
- Husband, R. W. (1931). Analysis of Methods in Human Maze Learning. *Pedagogical Seminary and Journal of Genetic Psychology*, 39, 258.
- Ishihara, S., & Force, R. A. A. (1943). *Ishihara Tests for Colour Blindness*: Sydney: Shephard & Newman.
- Jin, S.-A. A. (2009). Modality effects in Second Life: the mediating role of social presence and the moderating role of product involvement. *CyberPsychology & Behavior*, 12(6), 717-721.
- Johnson, W. L., & Rickel, J. (1997). Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environments. *Sigart Bulletin*, 8(1-4), 12-16.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments *International Journal of Artificial Intelligence in Education*, 11, 47-78.
- Johnson, W. L., Shaw, E., Marshall, A., & LaBore, C. (2003). *Evolution of User Interaction: The Case of Agent Adele*. Paper presented at the 8th International Conference on Intelligent user interfaces, January 12-15, 2003, Miami, Florida, USA.
- Karremans, J. C., Verwijmeren, T., Pronk, T. M., & Reitsma, M. (2009). Interacting with women can impair men's cognitive functioning. *Journal of Experimental Social Psychology*, 45(4), 1041-1044.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22-63.
- .

- Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1), 241-250.
- Kraemer, P. (2013). Do As We Do, Not As You Think: The Effect of Group Influence on Individual Choices in a Virtual Environment. *Journal For Virtual Worlds Research*, 6(1).
- Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction *Human-Computer Interaction: The Agency Perspective* (pp. 215-240): Springer.
- Kulms, P., Krämer, N. C., Gratch, J., & Kang, S.-H. (2011). It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze. *Intelligent Virtual Agents. Lecture Notes in Computer Science*, 6895, 80-92.
- Lombard, M. (2000). Resources for the study of presence: Presence explication. Retrieved from <http://nimbus.temple.edu/~mlombard/Presence/explicat.htm>
- Mason, R. (1994). *Using communications media in open and flexible learning*: Psychology Press.
- McCarthy, J. D., Sasse, M. A., & Riegelsberger, J. (2003). *Could I Have the Menu Please? An Eye Tracking Study of Design Conventions*. Paper presented at the HCI2003, 8-12 Sept, Bath, UK.
- McCroskey, J. C., Richmond, V. P., & Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research*, 1(4), 323-332.
- Monk, A. F., & Gale, C. (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3), 257-278.
- Mulac, A., Studley, L. B., Wiemann, J. M., & Bradac, J. J. (1987). Male/Female Gaze in Same-Sex and Mixed-Sex Dyads Gender-Linked Differences and Mutual Influence. *Human communication research*, 13(3), 323-343.
- Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). *A storytelling robot: Modeling and evaluation of human-like gaze behavior*. Paper presented at the 2006 6th IEEE-RAS International Conference on Humanoid Robots.
- Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology*, 29(5), 1093-1110.
- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *Int. Journal of Human-Computer Studies*, 43(2), 223-239. doi:<http://dx.doi.org/10.1006/ijhc.1995.1042>
- Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). *Anthropomorphism, agency, and ethopoeia: computers as social actors*. Paper presented at the INTERACT'93 and CHI'93 conference companion on Human factors in computing systems.
- Nauts, S., Metzmacher, M., Verwijmeren, T., Rommeswinkel, V., & Karremans, J. C. (2012). The mere anticipation of an interaction with a woman can impair men's cognitive performance. *Archives of sexual behavior*, 41(4), 1051-1056.
- Nielsen, G. S. (1963). *Studies in self confrontation : viewing a sound motion picture of self and another person in a stressful dyadic interaction / by Gerhard Nielsen*: Copenhagen : Munksgaard, [1963].

- Nowak, K. (2001). *Defining and differentiating copresence, social presence and presence as transportation*. Paper presented at the Presence, May 2001, Philadelphia, PA.
- Palmer, M. T. (1995). Interpersonal communication and virtual reality: Mediating interpersonal relationships. *Communication in the age of virtual reality*, 277-299.
- Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). *The Determinants of Web Page Viewing Behaviour: An Eye-Tracking Study*. . Paper presented at the 2004 Symposium on Eye Tracking Research and Applications, March 22-24 2004, San Antonio, Texas.
- Pessin, J. (1933). The Comparative Effects of Social and Mechanical Stimulation on Memorizing. *American Journal of Psychology*, 45, 263.
- Prendinger, H., Ma, C., Yingzi, J., Nakasone, A., & Ishizuka, M. (2005). *Understanding the Effect of Life-Like Interface Agents through Users' Eye Movements*. . Paper presented at the Seventh International Conference on Multimodal Interfaces (ICMI-05).
- Reid, E. (1994). Virtual Worlds: Culture and Imagination. *Cybersociety: Computer-mediated communication and community*, 164.
- Rickel, J., & Johnson, W. L. (1998). *Steve: A Pedagogical Agent for Virtual Reality*. Paper presented at the Proceedings of the Second International Conference on Autonomous Agents, Minneapolis / St. Paul.
- Savicki, V., & Kelley, M. (2000). Computer mediated communication: Gender and group composition. *CyberPsychology & Behavior*, 3(5), 817-826.
- Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375-397.
- Shahid, S., Krahmer, E., & Swerts, M. (2012). Video-mediated and co-present gameplay: Effects of mutual gaze on game experience, expressiveness and perceived social presence. *Interacting with Computers*, 24, 292-305.
- Shaw, E., Johnson, W. L., & Ganesan, R. (1999). *Pedagogical Agents on the Web*. Paper presented at the The Third International Conference on Autonomous Agents. ACM Press.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London ; New York: Wiley.
- Stephenson, G. M., & Rutter, D. R. (1970). Eye-contact, distance and affiliation: a re-evaluation. *British Journal of Psychology*, 61(3), 385.
- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American journal of psychology*, 9(4), 507-533.
- Vickers, S., Istance, H. O., Hyrskykari, A., Ali, N., & Bates, R. (2008). *Keeping an eye on the game: eye gaze interaction with Massively Multiplayer Online Games and virtual communities for motor impaired users*. Paper presented at the 7th International conference on disability, virtual reality and associated technologies;ICDVRAT; Sept. 8-11, 2008, , Maia, Portugal.
- Von Der Pütten, A. M., Klatt, J., Ten Broeke, S., McCall, R., Krämer, N. C., Wetzel, R., . . . Klatt, J. (2012). Subjective and behavioral presence measurement and interactivity in the collaborative augmented reality game TimeWarp. *Interacting with computers*, 24(4), 317-325.

- Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition*, 12(1), 85-101.
- Wang, N., & Gratch, J. (2010). *Don't just stare at me!* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA.
- Wiener, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in verbal communication*: Ardent Media.
- Witkowski, M., Arafa, Y., & de Bruijn, O. (2001). *Evaluating user reaction to character agent mediated displays using eye-tracking technology*. Paper presented at the AISB-01 Symposium on Information Agents for Electronic Commerce,.
- Witmer, B., & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3), 225-240.
- Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior*, 10(1), 115-121.
- Yuen, E. K., Herbert, J. D., Forman, E. M., Goetter, E. M., Comer, R., & Bradley, J.-C. (2013). Treatment of social anxiety disorder using online virtual environments in second life. *Behavior therapy*, 44(1), 51-61.
- Zajonc, R. B. (1965). Social Facilitation. *Science. New Series*, 149(3681), 269-274.
- Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., & Hodges, L. F. (2007). *Social responses to virtual humans: implications for future interface design*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

## 9 Appendices

### ***Appendix A: Consent form for Experiment 1***

You are invited to participate in a research study in which you will work with your partner to solve tasks within Second Life, an online virtual environment. This will involve you following instructions at a computer terminal. Your eye movements will be recorded whilst you are carrying out these tasks.

Please feel free to ask the experimenter any questions before the experiment begins. If, for any reason, you feel uncomfortable at any time during the experiment, please let the experimenter know.

The whole experiment should take between 30 and 45 minutes.

If you have read this form and have decided to participate in this project, please understand that your participation is voluntary and that you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study.

---

If you agree with the above-stated conditions and are willing to participate in the experiment, please sign below

Signed: \_\_\_\_\_ Date: \_\_\_\_\_



## ***Appendix B: Questionnaire for Experiment 1***

Age:

Under 18      18-24      25-30      31-40      41-50      Over 51

Gender:

Male

Female

Native language (please specify): .....

Were you giving instructions or following instructions?

Giving

Following

Please circle the number that you think most applies to you.

Overall, how did you enjoy doing the experiment?

1                      2                      3                      4                      5

(Not at all)

(Very Much)

Overall, how difficult did you find the experiment?

1                      2                      3                      4                      5

(Very easy)

(Very difficult)

How difficult did you find the first block of tasks?

1

2

3

4

5

(Very easy)

(Very difficult)

How difficult did you find the second block of tasks?

1

2

3

4

5

(Very easy)

(Very difficult)

Have you used Second Life before?

Never

1-2 times

3-10 times

more than 10 times

Do you have any comments about the overall experiment or suggestions about how it could be improved?

### ***Appendix C: Instructions for Agent condition (Experiment 2)***

Thank you for agreeing to take part in this study. This study compares 4 different programs for generating instructions. You will be interacting with Tracker, a computer-controlled agent – an Instruction Giver – in Second Life, an online virtual environment, in order to complete arithmetic problems. You will be given instructions by the agent, and you will select answers by clicking on one of the answer tiles at the bottom of the screen in front of you. The 4 programs for controlling the agents will have the same voice, and they will be unable to hear you, although they will be able to respond to what you are doing.

Please complete the tasks as quickly as possible; the 5 participants who complete the tasks in the shortest amount of time will each receive an Amazon voucher.

You will participate with each of the 4 programs, and will do 10 tasks with each agent. After interacting with each agent you will be asked some questions before interacting with the next agent.

### ***Appendix D: Instructions for Avatar condition (Experiment 2)***

Thank you for agreeing to take part in this study. This study compares 4 different programs for translating human behaviour into agent behaviour. You will be interacting with another person – an Instruction Giver – through Tracker, an avatar in Second Life, an online virtual environment, in order to complete arithmetic problems. You will be given instructions by your partner, who is located in a different room, and you will select answers by clicking on one of the answer tiles at the bottom of the screen in front of you. Your partner's voice will be disguised, and they will be unable to hear you, although they can see what you are doing and will be able to respond accordingly by controlling all of their avatar's movements and actions.

You will participate under each of the 4 programs, and will do 10 tasks with each program. After interacting with each program you will be asked some questions before interacting with the next program.

Please complete the tasks as quickly as possible; the 5 participants who complete the tasks in the shortest amount of time will each receive an Amazon voucher.

### ***Appendix E: Questionnaire items (Experiment 2)***

Q1	How much were you able to control Tracker's behaviour?
Q2	How responsive was Tracker to the actions you performed?
Q3	How completely were <i>all</i> your senses engaged in your interaction with Tracker?
Q4	How much did the appearance and behaviour of Tracker involve you in the interaction?
Q5	How much did Tracker's verbal instructions involve you in the interaction?
Q6	How aware were you of events occurring in the real world around you?
Q7	How aware were you of your display and control devices (mouse and monitor)?
Q8	How much did your interaction with Tracker seem consistent with your real-world interactions?
Q9	Were you able to anticipate what would happen next in response to the actions that you performed?
Q10	How clearly could you understand Tracker's instructions?
Q11	To what degree did you feel confused or disorientated at the beginning of breaks or at the end of the experimental session?
Q12	How involved were you in the experience of interacting with Tracker?
Q13	How distracting was the control mechanism?
Q14	How quickly did you adjust to the virtual environment experience?
Q15	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
Q16	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
Q17	How much did the control devices interfere with the performance of assigned tasks or with other activities?
Q18	How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
Q19	Did you learn new techniques that enabled you to improve your performance?
Q20	Were you involved in the experimental task to the extent that you lost track of time?

## ***Appendix F: Consent Form – Agent (Experiment 2)***

You are invited to participate in a research study in which you will work with a computer-controlled avatar to solve tasks within Second Life, an online virtual environment. This will involve you following instructions at a computer terminal. Your eye movements will be recorded whilst you are carrying out these tasks.

Please feel free to ask the experimenter any questions before the experiment begins. If, for any reason, you feel uncomfortable at any time during the experiment, please let the experimenter know.

The whole experiment should take no more than 1 hour.

If you have read this form and have decided to participate in this project, please understand that your participation is voluntary and that you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study.

---

If you agree with the above-stated conditions and are willing to participate in the experiment, please sign below

Signed: \_\_\_\_\_ Date: \_\_\_\_\_

## ***Appendix G: Consent Form – Avatar (Experiment 2)***

You are invited to participate in a research study in which you will work with a person in another room to solve tasks within Second Life, an online virtual environment. This will involve you following instructions at a computer terminal. Your eye movements will be recorded whilst you are carrying out these tasks.

Please feel free to ask the experimenter any questions before the experiment begins. If, for any reason, you feel uncomfortable at any time during the experiment, please let the experimenter know.

The whole experiment should take no more than 1 hour.

If you have read this form and have decided to participate in this project, please understand that your participation is voluntary and that you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study.

---

If you agree with the above-stated conditions and are willing to participate in the experiment, please sign below

Signed: \_\_\_\_\_ Date: \_\_\_\_\_

## ***Appendix F: Script 10c.50 (Experiment 2)***

```
// Use some textures
set-texture stimulus_tile1 = 6gc
set-texture stimulus_tile2 = 9bc
set-texture stimulus_tile3 = 1rs
set-texture stimulus_tile4 = 8gd
set-texture stimulus_tile5 = 3gs
set-texture stimulus_tile6 = 4bs
set-texture stimulus_tile7 = 7rd
set-texture response_tile1 = 31
set-texture response_tile2 = 32
set-texture response_tile3 = 36
//wait
sleep 5
// give some instructions and look at tiles
look-at stimulus_tile7
sleep 0.25
play-sound-file reddiamond
script openclose
look-at user tile
sleep 0.50
look-at instruction tile
sleep 0.25
play-sound-file multiply
script openclose1
look-at stimulus_tile1
sleep 0.25
play-sound-file greencircle
script openclose
look-at user tile
sleep 0.50
look-at instruction tile
sleep 0.25
play-sound-file plus
script openclose1
look-at stimulus_tile5
sleep 0.25
play-sound-file greensquare
script openclose
look-at user tile
sleep 0.50
look-at instruction tile
```

```
sleep 0.25
play-sound-file minus
script openclose1
look-at stimulus_tile2
sleep 0.25
play-sound-file bluecircle
script openclose
look-at user tile
sleep 0.50
look-at instruction tile
sleep 0.25
wait-for chat-from response_tile.*
// End of script script10c.50
//script script11a.50
```